

Indicadores bibliométricos de desempenho científico: estudo da aplicação de indicadores na avaliação individual do desempenho científico

Elizabeth de Sousa Vieira

Tese submetida à Faculdade de Engenharia da universidade do Porto para a obtenção do grau de doutor em Engenharia Industrial e Gestão

Orientadores

José Alberto Nunes Ferreira Gomes

José António Sarsfield Pereira Cabral



Universidade do Porto
Faculdade de Engenharia
FEUP

2013

Ao meu avô...

Resumo

Os indicadores bibliométricos têm sido utilizados, intensivamente, ao longo dos últimos anos para descrever o desempenho científico. O desenvolvimento destes indicadores tem recebido uma grande atenção porque as autoridades nacionais, instituições de financiamento público ou líderes de instituições que praticam investigação têm mostrado interesse em formas simples para descrever o desempenho científico das instituições. São vários os países que utilizam um sistema baseado na avaliação por pares e indicadores bibliométricos para avaliar o desempenho das instituições de ensino superior e posteriormente atribuir financiamento para as atividades de investigação.

A avaliação por pares continua a ser a principal metodologia utilizada quando se pretende avaliar o desempenho científico, mas a pressão para a realização de avaliações mais extensas e mais frequentes de instituições de investigação e de investigadores torna os indicadores bibliométricos um instrumento atrativo.

Neste trabalho estudou-se a aplicabilidade de indicadores bibliométricos como instrumento auxiliar das avaliações por pares em concursos para Professor Associado e Catedrático nas universidades portuguesas. O trabalho efetuado a este nível é praticamente inexistente e os resultados obtidos contribuem pouco para inferir sobre o papel que os indicadores de desempenho podem ter na formulação das decisões finais na avaliação por pares.

Nos concursos são vários os aspetos avaliados: desempenho pedagógico, desempenho científico, atividades de gestão e transferência de tecnologia. Apenas as variáveis que descrevem o desempenho científico (neste caso indicadores bibliométricos) foram consideradas. Para os vários candidatos de cada concurso foi determinado um conjunto de indicadores que se acredita descreverem o desempenho científico do respetivo investigador e que poderão estar implícitos nas decisões dos painéis de avaliação. Posteriormente desenvolveram-se modelos que pretendem traduzir os julgamentos finais de um dado painel. Os modelos obtidos foram testados relativamente ao seu poder de previsão com o objetivo de determinar o grau de aderência das previsões baseadas em modelos que usam indicadores à realidade das decisões finais nos concursos.

Com este trabalho não se pretende sugerir que em concursos deste tipo as avaliações devem ser exclusivamente baseadas nos resultados fornecidos pelos indicadores bibliométricos, pois existem vários aspetos que apenas podem ser aferidos utilizando o conhecimento e experiência do ser humano.

Os resultados obtidos mostraram que modelos baseados em indicadores bibliométricos podem, de facto, ser utilizados como instrumento auxiliar contribuindo para a formulação das decisões finais e eliminando algumas das limitações inerentes à avaliação por pares. Com a realização desta tese demonstrou-se o potencial dos indicadores bibliométricos na avaliação do desempenho científico de indivíduos.

Abstract

Bibliometric indicators have been widely used for assessing the scientific performance of a given research body. The design of indicators has attracted a lot of attention in the last few years as national authorities, funding bodies or institutional leaders show a growing interest in indicators that can rate the performance of their institutions. Nowadays, several countries use a combination of peer review and bibliometric indicators to assess the research performance of higher education institutions and to allocate funding.

Peer review is still the gold standard of research evaluation but the pressure for more frequent and extensive assessment of the performance of researchers, research groups and research institutions makes bibliometry attractive.

In this study was explored the use of bibliometric indicators as auxiliary instrument of peer review at academic openings for Professor *Associado* and *Catedrático* at Portuguese universities. The studies that have been done at this level are just a few and the results obtained are not enough to realize the role that bibliometric indicators can play in formulating the final judgments in the peer review process.

In these academic openings several dimensions are assessed: pedagogical activities, management activities, knowledge transfer and scientific performance. Only those variables (bibliometric indicators) that describe the scientific performance were taken into account. A set of bibliometric indicators, which it is believed, to describe the individual performance and that are implicit in peer decisions were calculated for each applicant. With the bibliometric indicators available were developed several models that aim at describing the final judgments. The predictions given by the models were studied in detail in order to identify how far the models can predict the peer decisions on the academic openings.

The study outlined in this thesis does not pretend to suggest that in this type of academic opening the final decisions should be based on the results given by the bibliometric indicators, exclusively. There are some aspects of the scientific performance that can be only assessed using the human experience and knowledge.

The results obtained showed that models based on bibliometric indicators can be used as auxiliary instrument in peer review at academic openings. These models can help peers in the evaluation and formulation of the final decisions avoiding some of the limitations associated with peer review. The study showed the potential of bibliometric indicators on the evaluation of individual scientific performance.

Agradecimentos

O meu primeiro obrigado vai para o Professor José Alberto Nunes Ferreira Gomes (orientador) e José António Sarsfield Pereira Cabral (co-orientador) pelo trabalho de orientação prestado ao longo de todo o desenvolvimento da minha tese. Obrigada pelo tempo despendido e conhecimento partilhado sem o qual não seria possível a realização do trabalho aqui apresentado.

Também agradeço à Fundação para a Ciência e Tecnologia (FCT) pela bolsa (SFRH / BD / 75190 / 2010) cofinanciada pelo QREN-POPH Tipologia 4.1 – Formação avançada, comparticipada pelo Fundo Social Europeu e por fundos nacionais do MEC.

Acrónimos

<i>A&HCI</i>	<i>Arts & Humanities Citation Index</i>
<i>AIC</i>	Critério de Informação de Akaike
<i>APD</i>	Número de autores por documento
<i>ARC</i>	<i>Australian Research Council</i>
<i>BIF</i>	<i>Boehringer Ingelheim Fonds</i>
<i>BKCI-S</i>	<i>Book Citation Index – Science</i>
<i>BKCI-SSH</i>	<i>Book Citation Index – Social Sciences & Humanities</i>
<i>CA</i>	<i>Chemistry, Applied</i>
<i>CD</i>	<i>Citing documents</i>
<i>CI</i>	Índice de condição
<i>CICECO</i>	Centro de Investigação em Materiais Cerâmicos e Compósitos
<i>CM</i>	<i>Chemistry, Multidisciplinary</i>
<i>CO</i>	<i>Chemistry, Organic</i>
<i>CP</i>	<i>Chemistry, Physical</i>
<i>CPCI-S</i>	<i>Conference Proceedings Citation Index - Science</i>
<i>CPCI-SSH</i>	<i>Conference Proceedings Citation Index - Social Science & Humanities</i>
<i>CPD</i>	Média de citações por documento
<i>CPP/FCSm</i>	<i>Crown indicator</i>
<i>CRUP</i>	Concelho de Reitores de Universidades Portuguesas
<i>CSIC</i>	<i>Consejo Superior de Investigaciones Científicas</i>
<i>CWTS</i>	<i>Centre for Science and Technology Studies</i>
<i>DGEEC</i>	Direção Geral de Estatísticas da Educação e Ciência
<i>DIC</i>	<i>Documents with international collaboration</i>
<i>EC</i>	<i>Engineering, Chemical</i>
<i>ERA</i>	<i>Excellence in Research for Australia</i>
<i>ESI</i>	<i>Essential Science Indicators</i>
<i>HCD</i>	<i>Highly cited documents</i>
<i>HSF</i>	<i>Heart Stroke Foundation</i>
<i>IBMC</i>	Instituto de Biologia Molecular e Celular
<i>IIA</i>	Independência de Alternativas Irrelevantes
<i>INESC Porto</i>	Instituto de Engenharia de Sistemas e Computadores do Porto
<i>JCR</i>	<i>Journal Citation Reports</i>
<i>JRI</i>	<i>Journal Relative Impact</i>
<i>KMO</i>	<i>Kaiser-Meyer-Olkin</i>
<i>MA</i>	<i>Mathematics, Applied</i>
<i>MSA</i>	Medida de Adequação da Amostra
<i>MRC</i>	<i>Medical Research Council of Canada</i>
<i>MSM</i>	<i>Materials Science, Multidisciplinary</i>
<i>NA_m</i>	<i>Number of Authors per Document</i>
<i>NDF</i>	<i>Number of Documents Fractioned</i>
<i>NI</i>	<i>Normalized Impact</i>
<i>NIR</i>	<i>Normalized Indicator for Researchers</i>
<i>PCD</i>	<i>Percentage of Cited Documents</i>

<i>PP</i>	<i>Pharmacology & Pharmacy</i>
<i>QI</i>	<i>High Quality Impact</i>
<i>RAE</i>	<i>Research Assessment Exercise</i>
<i>REF</i>	<i>British Research Excellence Framework</i>
<i>REQUIMTE</i>	<i>Rede de Química e Tecnologia</i>
<i>ROLR</i>	<i>Rank-Ordered Logistic Regression</i>
<i>SC</i>	<i>Categorias do Web of Science</i>
<i>SCI-EXPANDED</i>	<i>Science Citation Index Expanded</i>
<i>SIR</i>	<i>SCImago Institutions Ranking</i>
<i>SJA</i>	<i>Scopus Journal Analyser</i>
<i>SJR</i>	<i>SCImago Journal Rank</i>
<i>SNIP</i>	<i>Source Normalized Impact per Paper</i>
<i>SSCI</i>	<i>Social Sciences Citation Index</i>
<i>TD</i>	<i>Número total de documentos</i>
<i>TD h-core</i>	<i>Percentagem do total de documentos que definem o índice h</i>
<i>TD h_{nf}-core</i>	<i>Percentagem do total de documentos que definem o índice h_{nf}</i>
<i>UKAIS</i>	<i>UK Academic for Information System</i>
<i>VIF</i>	<i>Fator de inflação da variância</i>
<i>VQR</i>	<i>Valutazione Quinquennale della Ricerca</i>
<i>VSNU</i>	<i>Association of Universities in the Netherlands</i>
<i>VTR</i>	<i>Valutazione Triennale della Ricerca</i>
<i>WoS</i>	<i>Web of Science</i>
w_i	<i>Peso de Akaike</i>
ρ	<i>Índice da razão da máxima verosimilhança</i>

Índice

CAPÍTULO 1	1
Introdução	1
1.1 O contexto da análise bibliométrica	3
1.1.1 Os “pilares” da análise bibliométrica	4
1.1.2 A influência da cultura de publicação e citação	5
1.2 Bases de referência bibliográfica	7
1.2.1 Thomson Reuters Web of Knowledge	9
1.2.2 SciVerse Scopus	12
1.2.3 As limitações do Scopus e WoS na indexação	13
1.2.4 Recolha da informação no Scopus e WoS	15
1.3 Relação entre indicadores bibliométricos e a avaliação por pares	17
1.4 O uso de análises bibliométricas em Portugal	22
1.5 Motivação e objetivos da investigação	23
1.6 Delineamento da tese	25
 CAPÍTULO 2	 27
Estudo das bases multidisciplinares	27
2.1 Amostra	30
2.2 Metodologia	31
2.2.1 Pesquisa da produção científica no Scopus e WoS para instituições	31
2.2.2 Pesquisa da produção científica no WoS para investigadores	34
2.2.3 Quantificação da produção científica	39
2.2.4 Quantificação do impacto da produção científica	39
2.2.5 Análise das bases de citações do WoS	39
2.3 Documentos indexados no WoS e/ou Scopus para uma instituição	41
2.4 O impacto no Scopus e/ou WoS para uma instituição	46
2.5 Bases de citações que compõem o WoS	48
2.6 Conclusões	51
 CAPÍTULO 3	 53
Indicadores bibliométricos	53
3.1 Amostra	55
3.2 Metodologia	56
3.2.1 Cálculo dos indicadores	56
3.2.2 Metodologia de pesquisa no WoS para investigadores	68
3.3 Indicadores desenvolvidos	69
3.3.1 Impacto relativo do periódico (<i>Journal Relative Impact, JRI</i>)	69
3.3.2 O índice h_{nf}	89
3.3.3 Impacto normalizado para investigadores (<i>Normalized Indicators for Researchers, NIR</i>)	102
3.3.4 Número de autores por documento (<i>Number of Authors per Document, NA_m</i>)	109
3.4 Conclusões	115

CAPÍTULO 4	117
Definição do Modelo	117
4.1 Amostra	119
4.2 Metodologia	120
4.2.1 Pesquisa da produção científica no <i>WoS</i> para investigadores	120
4.2.2 Análise estatística	120
4.3 Definição do modelo	135
4.3.1 Análise Fatorial	135
4.3.2 <i>Rank-ordered logit regression</i>	141
4.4 Conclusões	151
CAPÍTULO 5	153
Poder de previsão do modelo	153
5.1 Amostra	155
5.2 Metodologia	155
5.3 O sucesso dos modelos na previsão dos resultados	164
5.4 Função de distribuição de probabilidades	165
5.5 Comparação das previsões dos modelos com um cenário aleatório	168
5.6 Poder de previsão dos modelos quando apenas a produção científica é avaliada	169
5.7 Conclusões	173
CAPÍTULO 6	175
Conclusões gerais	175
CAPÍTULO 7	181
Lista de publicações	181
CAPÍTULO 8	185
Referências	185

Lista de Figuras

Figura 1. Metodologia de pesquisa utilizada na obtenção da produção científica indexada no <i>Scopus</i> e <i>WoS</i> das universidades do CRUP.....	32
Figura 2. Metodologia de pesquisa utilizada na obtenção da produção científica indexada no <i>WoS</i> dos investigadores.	35
Figura 3. A afiliação de cada autor do documento e restantes aspetos que caracterizam uma publicação indexada no <i>WoS</i>	38
Figura 4. Número de documentos indexados no ano 2006 em apenas uma das bases de dados e respetiva média de citações por documento (impacto).	48
Figura 5. Evolução do número médio de citações por documento (<i>CPD</i>) em <i>Chemistry</i> , <i>Physical</i> e <i>Statistic & Probability</i> para os documentos publicados entre 1990 e 2008.	72
Figura 6. Influência das características da janela de citação na previsão do número médio de citações por documento (<i>CPD</i>) para um conjunto periódicos classificados na categoria <i>Statistics & Probability</i>	75
Figura 7. Influência das características da janela de citação na previsão do número médio de citações por documento (<i>CPD</i>) para um conjunto de periódicos classificados na categoria <i>Chemistry</i> , <i>Physical</i>	76
Figura 8. Valores de <i>JRI</i> , <i>SJR</i> e <i>SNIP</i> para um conjunto de periódicos.	83
Figura 9. Diferenças observadas nas ordenações para o <i>JRI</i> , <i>SNIP</i> e <i>SJR</i>	84
Figura 10. Relação entre a probabilidade determinada e a componente determinística da utilidade (Train 2009).	131
Figura 11. Índice da razão de máxima verosimilhança para as diferentes variáveis dependentes.	144
Figura 12. Correlações obtidas entre as probabilidades determinadas com a metodologia <i>jackknife</i> e as probabilidades iniciais.	146
Figura 13. Evolução da componente determinística com o valor assumido em cada percentil pelos indicadores que definem os modelos.	149
Figura 14. Contagem de pares para um concurso com seis candidatos e cuja ordenação é a conhecida.	156
Figura 15. Histograma de distribuição de probabilidades de P_{ni}	165
Figura 16. Função distribuição de probabilidades obtidas para cada modelo.	166
Figura 17. Ajuste da distribuição Normal a cada um dos modelos.	167
Figura 18. Relação entre a probabilidade de um candidato ser selecionado em primeiro se apenas o desempenho científico é avaliado e a probabilidade para o caso real onde a avaliação é mista e diferentes pesos são atribuídos às dimensões avaliadas.	170
Figura 19. Função distribuição de probabilidades para cada modelo considerando que apenas o desempenho científico é avaliado.	171

Lista de tabelas

Tabela 1.	Número de documentos indexados no <i>WoS</i> e <i>Scopus</i> para cada uma das universidades do CRUP.	41
Tabela 2.	Número de documentos indexados no <i>WoS</i> e <i>Scopus</i> para cada uma das universidades do CRUP (continuação).	42
Tabela 3.	Número de documentos indexados no <i>Scopus</i> e <i>WoS</i> para a Universidade de Coimbra e Universidade de Lisboa em 2006.	43
Tabela 4.	Justificação para a ausência de alguns documentos em uma ou outra base de dados em 2006.	45
Tabela 5.	Documentos indexados em 2006 para a Universidade de Coimbra e Universidade de Lisboa, após a correção para possíveis erros na indexação (B1) e alterações da estrutura da base de dados (A2).	45
Tabela 6.	Distribuição do número de documentos indexados no <i>Scopus</i> por área científica para a Universidade de Coimbra e Universidade de Lisboa em 2006.	46
Tabela 7.	A média de citações por documento (<i>CPD</i>) obtido para cada um dos conjuntos de documentos indexados em 2006.	47
Tabela 8.	Número de documentos indexados nas várias bases de citação do <i>WoS</i> para o conjunto de 210 investigadores holandeses no período 2003-2007.	49
Tabela 9.	Número de citações (<i>TC</i>) e número médio de citações por documento (<i>CPD</i>) contabilizadas nas várias bases de citação do <i>WoS</i> para as categorias consideradas, no período 2003-2007 (amostra de 210 investigadores holandeses).	50
Tabela 10.	Exemplo de cálculo do h_{nf}	65
Tabela 11.	Número médio de citações por documento (<i>CPD</i>) para diferentes tipos de documentos e categorias.	69
Tabela 12.	Contabilização das citações utilizando a metodologia diacrónica.	73
Tabela 13.	Contabilização das citações utilizando a metodologia síncrona.	73
Tabela 14.	Comprimento da janela de citação para S1 e S2.	77
Tabela 15.	Média, desvio padrão e coeficiente de variação para o <i>JRI</i> , <i>SNIP</i> e <i>SJR</i> determinados para cinco períodos.	80
Tabela 16.	Média, desvio padrão e coeficiente de variação para o <i>JRI</i> , <i>SNIP</i> e <i>SJR</i> determinados para cinco períodos (continuação).	81
Tabela 17.	Valores determinados para o índice h e h_f na <i>PP</i>	90
Tabela 18.	Valores determinados para o índice h e h_f na <i>MA</i>	91
Tabela 19.	Valores obtidos para o índice h e h_{nf} na <i>PP</i>	91
Tabela 20.	Valores obtidos para o índice h e h_{nf} na <i>MA</i>	92
Tabela 21.	Média, desvio padrão e coeficiente de variação para diferentes indicadores.	93
Tabela 22.	Ordenação decrescente dos investigadores utilizando os valores de h_{nf}	94
Tabela 23.	Valores de h_{nf} , h , <i>TC</i> e <i>TD</i> em 1996-2003 e 2000-2007 para a <i>PP</i>	95
Tabela 24.	Ordenação decrescente dos investigadores para cada indicador.	98
Tabela 25.	Cálculo do índice g	99
Tabela 26.	Valores obtidos para o <i>CPD</i> e <i>NIR</i> na <i>PP</i>	103
Tabela 27.	Valores obtidos para o <i>CPD</i> e <i>NIR</i> na <i>MA</i>	104
Tabela 28.	Ordenação dos investigadores utilizando o <i>CPD</i> e o <i>NIR</i>	105
Tabela 29.	Valores para o <i>TD</i> , <i>CPD</i> e o <i>NIR</i> em 1996-2003 e 2000-2007 na <i>PP</i>	106

Tabela 30.	Informação utilizada no processo de normalização do <i>CPP/FCSm</i>	107
Tabela 31.	Dados utilizados no cálculo do <i>MNCS</i> e <i>NIR</i>	108
Tabela 32.	Número médio de autores por documento (<i>APD</i>) para diferentes tipos de documentos e categorias.	110
Tabela 33.	Valores obtidos para o <i>APD</i> e <i>NA_m</i> na <i>PP</i>	111
Tabela 34.	Valores obtidos para o <i>APD</i> e <i>NA_m</i> na <i>MA</i>	111
Tabela 35.	Ordenação dos investigadores utilizando o <i>APD</i> e o <i>NA_m</i>	112
Tabela 36.	Valores para o <i>TD</i> , <i>APD</i> e <i>NA_m</i> em 1996-2003 e 2000-2007 na <i>PP</i>	113
Tabela 37.	Valores obtidos para o <i>VIF</i>	135
Tabela 38.	Estatísticas descritivas para a amostra de investigadores.	136
Tabela 39.	Valores obtidos no teste de <i>Shapiro-Wilk</i>	136
Tabela 40.	Matriz de correlações para os 12 indicadores (variáveis) utilizados.	137
Tabela 41.	Matriz anti-imagem.	137
Tabela 42.	Proporção da variância total explicada por cada fator após a rotação.	139
Tabela 43.	Pesos fatoriais e comunalidades obtidos na análise fatorial.	139
Tabela 44.	Valores obtidos para o <i>Cronbach's α</i>	140
Tabela 45.	Valores de <i>VIF</i> para as variáveis e do <i>CI</i>	141
Tabela 46.	Distribuição dos candidatos nos 27 concursos.	142
Tabela 47.	Variáveis com impacto significativo ($p < 0,05$) retidas quando são utilizadas as diferentes variáveis dependentes.	143
Tabela 48.	Parâmetros estimados para cada modelo.	145
Tabela 49.	Valores obtidos para <i>AIC</i> e o w_i	147
Tabela 50.	Valores determinados para a componente determinística.	148
Tabela 51.	Valores determinados para as elasticidades para o valor médio.	149
Tabela 52.	Valores determinados para as elasticidades para o valor mediano.	149
Tabela 53.	Comparação dos modelos com um cenário aleatório.	164
Tabela 54.	Valores de simetria, curtose, média e desvio padrão obtidos para as distribuições de cada modelo.	167
Tabela 55.	Média e desvio padrão para cada função distribuição de probabilidade. ..	169
Tabela 56.	Valores estimados para a dispersão de <i>V2</i>	172

Capítulo 1

Introdução

1.1 O contexto da análise bibliométrica

A ciência não pode resolver o mistério final da natureza. E isto porque em última análise, somos parte do mistério que tentamos resolver.

(Max Planck)

A ciência representa a aquisição de conhecimento através da realização de experiências e observação de comportamentos que permitem explicar fenômenos naturais e sociais. Os investigadores estudam um conjunto de evidências acerca destes fenômenos na tentativa de construir teorias que expliquem os acontecimentos observados no passado, no presente e que ocorrerão no futuro.

A avaliação dos resultados da investigação científica é uma prática muito comum em todo o mundo. Isto permite identificar os pontos fortes e fracos de um país ao nível da ciência e fazer a distribuição do financiamento destinado à investigação de acordo com as metas de desempenho estabelecidas para o sistema científico.

Os resultados da ciência têm sido medidos, ao longo dos anos, utilizando indicadores económicos, indicadores sociais, a avaliação por pares e mais recentemente técnicas bibliométricas.

O termo bibliometria foi criado em 1934 por Paul Otlet, mas só após a publicação por *Pritchard* do documento intitulado “*Statistical bibliography or bibliometrics*” em 1969 (*Pritchard*, 1969) é que o termo começou a ser popular. Segundo *Pritchard* as técnicas bibliométricas resultam da aplicação de um conjunto de métodos estatísticos e matemáticos cujo objetivo passa por definir o processo de comunicação, a natureza e desenvolvimento das áreas científicas. Isto é conseguido partindo da contagem e análise dos vários aspetos que descrevem as atividades de investigação.

As técnicas bibliométricas tiveram uma grande evolução a partir dos anos 70 devido à necessidade de estudar o efeito das políticas implementadas em vários países nos sistemas científicos e à necessidade de identificar as tendências das atividades de investigação.

Um conjunto de indicadores bibliométricos que permitem descrever as várias dimensões que caracterizam a produção científica de um país, instituição de ensino superior, instituição de investigação e indivíduos são utilizados na bibliometria.

Atualmente existem vários *rankings* baseados unicamente num conjunto de indicadores bibliométricos determinados para instituições vocacionadas para as atividades de investigação e cujo objetivo é descrever o desempenho científico das mesmas. O *SCImago Institutions Ranking (SIR)* produzido pelo *SCImago* e o *Leiden Ranking* produzido pelo *Centre for Science and Technology Studies (CWTS)* são *rankings* deste tipo. O *Times Higher Education World University Rankings* produzido pela *Thomson Reuters* e que pretende avaliar o desempenho de instituições de ensino superior em vários aspetos, incluindo a investigação, também utiliza indicadores bibliométricos. O *Academic Ranking of World Universities* elaborado por *Shanghai Jiao Tong University* também avalia o desempenho de instituições de ensino superior, embora num menor número de aspetos que o *Times Higher Education World University Rankings*. Os indicadores bibliométricos entre outros indicadores são usados para este efeito.

Para além da aplicação na elaboração de *rankings*, as técnicas bibliométricas começam também a ser aplicadas em conjunto com a avaliação por pares na avaliação dos sistemas científicos de vários países. Exemplos de sistemas de avaliação mistos são o *Excellence in Research for Australia (ERA)* e o *Valutazione Quinquennale della Ricerca (VQR)*. O *British Research Excellence Framework (REF)*, no Reino Unido, que substituiu o *Research Assessment Exercise (RAE)* também permite aos pares a utilização de indicadores bibliométricos para a avaliação das submissões efetuadas por cada universidade. No entanto, a avaliação por pares é o principal meio de avaliação e os indicadores bibliométricos apenas podem ser utilizados como instrumento auxiliar. Na Holanda o sistema de avaliação científico estabelecido nos anos 90's pelo *Association of Universities in the Netherlands (VSNU)* e cujo objetivo é avaliar os departamentos das instituições de ensino superior, utiliza a avaliação por pares, mas em disciplinas como a física, química e biologia as análises bibliométricas têm um papel formal. Outros países como a Alemanha, Finlândia e Noruega começam também a utilizar metodologias de avaliação mistas (OECD 2010).

1.1.1 Os “pilares” da análise bibliométrica

A análise bibliométrica baseia-se, essencialmente, na contagem bibliográfica e na análise de citações.

Contagem bibliográfica

A contagem bibliográfica implica a utilização de indicadores que permitem quantificar os produtos finais das atividades de investigação. Normalmente estes indicadores são baseados no número de artigos, livros e outros tipos de documentos publicados. As patentes registadas são também uma medida utilizada para quantificar a produção.

Análise de citações

No desenvolvimento dos trabalhos científicos os investigadores baseiam-se, normalmente, nos resultados divulgados em documentos previamente publicados. Estes resultados influenciam o pensamento do investigador contribuindo para a resolução de problemas. Tais contributos são reconhecidos citando-se a investigação que serviu de base para atingir os objetivos do trabalho elaborado. Quando um dado trabalho recebe muita atenção da comunidade científica (medida pelo número de citações) isto é visto como um indicador da sua importância e visibilidade. É neste sentido que as citações constituem um instrumento que permite descrever o impacto das atividades de investigação.

As citações podem ser utilizadas para obter informação sobre o impacto das atividades de investigação de um indivíduo, departamento de investigação, instituições de investigação ou outros, mas não para inferir sobre a sua qualidade científica no global. A qualidade científica é um conceito multidimensional e que não pode ser caracterizado em toda a sua extensão utilizando apenas a análise de citações. No entanto, se uma agência de avaliação pretende fazer uso da análise de citações as dimensões que se pretendem avaliar devem ser enumeradas de forma clara. Com este procedimento os resultados da análise de citações

podem contribuir para a definição da qualidade do desempenho científico, em conjunto com outros aspetos.

O documento publicado por Gross e Gross 1927, constitui um dos primeiros trabalhos onde a análise de citações foi utilizada para avaliar o impacto de trabalhos científicos. Desde então são vários os estudos que têm sido desenvolvidos com o objetivo de mostrar as potencialidades da análise de citações na descrição do impacto da produção científica (exemplos: Costas e Bordons 2005; King 2004; Lewison 1998; Ventura e Mombru 2006).

1.1.2 A influência da cultura de publicação e citação

Como foi mencionado anteriormente a análise bibliométrica assenta essencialmente em dois pilares: a contagem bibliográfica e a análise de citações. Partindo disto, indicadores bibliométricos que permitem descrever as várias dimensões do desempenho científico podem ser elaborados. Na elaboração destes especial atenção deve ser dada às culturas de publicação e citação das diferentes áreas científicas pois apenas os indicadores que têm em consideração os diferentes hábitos de publicação e citação podem ser utilizados para comparar conjuntos de documentos de diferentes áreas.

Cultura de publicação

O meio de comunicação que é escolhido para a disseminação dos resultados de investigação não é o mesmo nas diversas áreas científicas. Nas ciências sociais por exemplo os investigadores publicam mais em livros do que nas ciências naturais (Hicks 1999; Kyvik 2003; Lariviere, Gingras e Archambault 2006). Nas ciências dos computadores e engenharias os investigadores publicam mais em *proceedings* das conferências do que em outras áreas (Butler 2008; Glanzel, Schlemmer, Schubert *et al.* 2006; Moed e Visser 2007; Montesi e Owen 2008).

O número de autores por documento é também algo característico de cada área científica. Por exemplo, para a química e física o número médio de autores por documento é superior ao da matemática (Franceschet e Costantini 2011; Vieira e Gomes 2010) devendo-se isto às diferentes práticas de colaboração que caracterizam cada área científica.

No momento de elaboração de indicadores cujo objetivo passa pela quantificação da produção científica estes aspetos não podem ser esquecidos.

Cultura de citação

No momento de definição de indicadores baseados na análise de citações os vários aspetos que caracterizam as culturas de citação devem ser, igualmente, considerados.

As citações são contabilizadas fazendo uso da lista de referências presente em cada documento publicado. A dimensão das listas de referência é variável de área para área (Vieira e Gomes 2010). Por exemplo para a matemática a média de referências por documento é inferior à média para a química. Isto demonstra que na matemática os investigadores têm por hábito citar menos e é por isso natural que os seus trabalhos recebam menos citações.

O tipo de documento deve ser tido em consideração na análise de citações. Existem documentos como por exemplo os artigos de revisão que em média recebem mais citações do que artigos ou *proceedings paper*. Não é só o número de citações recebidas que difere, o tempo de vida é também diferente. Como tempo de vida de um documento entenda-se o período em que o mesmo é citado por outros. O estudo realizado por Lisee, Lariviere e Archambault 2008 mostrou que nos *proceedings paper* um número considerável de citações é atingido num espaço de tempo inferior ao observado para os restantes tipos de documento e que os mesmos se tornam obsoletos mais rapidamente. Este comportamento é esperado uma vez que os *proceedings paper* constituem o principal meio de divulgação dos resultados preliminares das atividades de investigação.

A maturação das citações acontece para períodos diferentes conforme as áreas científicas. Nas áreas científicas caracterizadas por um elevado tráfego de citações a maturação das citações ocorre num período de tempo mais curto do que nas áreas onde o tráfego de citações é mais baixo. Por maturação das citações considera-se o período de tempo para o qual o pico máximo de citações acontece. O estudo efetuado por Glanzel e Schoepflin 1995 retrata detalhadamente a evolução das citações ao longo do tempo para diferentes áreas científicas e permite concluir que se deve ter um cuidado especial na definição da amplitude da janela que será utilizada na análise de citações.

Todos os aspetos apontados anteriormente indicam que a análise de citações não deve considerar indicadores tais como o número total de citações contabilizado para um dado período de tempo ou o número médio de citações sem estarem normalizados. Estes indicadores não permitem a comparação da produção científica originada nas diferentes áreas de investigação.

Ao longo dos anos são vários os indicadores normalizados que têm sido propostos para este fim (exemplos: Gonzalez-Pereira, Guerrero-Bote e Moya-Anegon 2010; Moed 2010; Moed, Debruin e van Leeuwen 1995; Schubert e Braun 1986; Waltman, van Eck, van Leeuwen *et al.* 2011b; Zitt e Small 2008).

Os processos de normalização mais utilizados baseiam-se, essencialmente, na comparação da média de citações de um conjunto de documentos com uma média de citações por documento que é estabelecida como referência. A média de citações por documento mais usada como referência é a determinada para o conjunto de documentos que representam a produção científica de todo o mundo. A área científica em que os documentos se inserem, o tipo de documento e o ano de publicação são tidos em consideração no processo de normalização. Note-se que a distribuição das citações se caracteriza por uma elevada assimetria (exemplos: Redner 1998; Seglen 1992; Vieira e Gomes 2010). Existem também processos de normalização que utilizam como referência outros parâmetros da distribuição de citações (Bornmann, Leydesdorff e Mutz 2013), mas a utilização do valor médio é sem dúvida o mais aplicado pela comunidade científica (exemplos: Glanzel, Thijs, Schubert *et al.* 2009; Moed, Debruin e van Leeuwen 1995; Schubert, Glanzel e Braun 1986; Waltman, van Eck, van Leeuwen *et al.* 2011a, 2011b).

O recurso à análise de citações é mais comum ao nível macro (países e instituições de ensino superior). Para estes níveis considera-se que o efeito negativo que poderá advir da utilização de valores médios que resultam de distribuições muito enviesadas é mínimo pelo elevado número de documentos que são considerados no cálculo dos indicadores normalizados. Quando a análise de citações é aplicada a departamentos ou a indivíduos deve

ser acompanhada de uma interpretação cuidadosa dos resultados obtidos dado o menor número de documentos utilizado.

Até este ponto foram mencionados os aspetos que devem ser considerados no desenvolvimento de indicadores utilizados na análise de citações. No entanto é importante ter em consideração que outros fatores podem influenciar as citações:

- A visibilidade internacional do meio de comunicação utilizado na divulgação dos resultados científicos (Vieira e Gomes 2010).
- O idioma utilizado na comunicação dos resultados. Para os documentos redigidos em inglês é esperado um maior número de citações.
- A disponibilidade das publicações é importante. Publicações cujo conteúdo pode ser consultado sem efetuar qualquer tipo de pagamento (*open access*) pela comunidade que utiliza a Internet terão maior probabilidade de ser citadas em algumas áreas científicas (Donovan e Watson 2011).

1.2 Bases de referência bibliográfica

As análises bibliométricas apenas poderão ser efetuadas se existirem instrumentos que fornecem a informação suficiente para este fim. O desenvolvimento de indicadores bibliométricos recorre a bases de dados de referência bibliográfica.

Atualmente são várias as bases de referência de documentos de índole científica. Estas bases de dados podem ser classificadas em dois grandes grupos: bases de dados multidisciplinares como o *Sciverse Scopus*, *Thomson Reuters Web of Knowledge* e o *Google Scholar* e bases de dados mais especializadas como por exemplo o *Engineering Village*, *MEDLINE*, *MathSciNet®*, *Chemical Abstracts Service (CAS)* e *Econlit*.

As bases de dados multidisciplinares são as mais utilizadas quando se pretende fazer análises bibliométricas. A oferta de um conjunto alargado de instrumentos de análise torna-as muito úteis para a elaboração de estudos bibliométricos. As bases de dados mais especializadas têm vindo a evoluir a este nível, mas ainda não apresentam as mesmas vantagens das bases de dados multidisciplinares.

Das três bases de dados multidisciplinares disponíveis apenas se apresenta uma breve descrição do *Sciverse Scopus* e do *Thomson Reuters Web of Knowledge*. A exclusão do *Google Scholar* foi baseada nos vários estudos que têm vindo a ser efetuados ao longo dos anos e cujo objetivo passou, na maioria dos casos, pela comparação das diversas características apresentadas pelas três bases de dados. Nestes estudos várias críticas foram apontadas ao *Google Scholar* como instrumento para análise bibliométrica. As principais críticas estão resumidas em seguida:

- Não existe informação acerca do número de registos indexados ou período de indexação dos documentos publicados nos mais diversos meios de comunicação.
- O controlo de qualidade dos diversos meios de comunicação dos resultados das atividades de investigação não existe. Existe sim a indexação de todo o tipo de material, incluindo, por exemplo, o disponível em repositórios *preprints*. Os repositórios de *preprints* devem ser tratados com muito cuidado pois contêm as

versões iniciais das publicações antes de serem publicadas oficialmente após o processo de avaliação por pares.

- Para um documento apenas é possível efetuar a pesquisa de informação utilizando um pequeno número de campos. A pesquisa por endereço, por exemplo, não está acessível.
- A exportação da informação para *software* de referências bibliográficas é muito primitiva, permitindo apenas a exportação de um documento de cada vez.
- Não oferece informação suficiente para a construção de indicadores normalizados, que são indispensáveis na análise de citações.
- Aparecem com muita frequência documentos duplicados.
- Vários problemas técnicos foram detetados. Jacso 2006 verificou que os operadores booleanos não funcionam corretamente. O algoritmo desenhado para a recolha das citações apresenta graves deficiências obrigando a uma correção manual da contagem de citações o que se pode tornar num processo bastante moroso (Jacso 2008; Meho e Yang 2007).
- No *Google Scholar* é possível a um investigador manipular os resultados providenciados pelo *Google Citations* através da criação de documentos falsos que citam os documentos do investigador (Delgado Lopez-Cozar, Robinson-Garcia e Torres Salinas 2012). Isto é possível porque nesta base de dados está disponível um conjunto de instrumentos que não estão submetidos a qualquer tipo de controlo.

Tendo em consideração as críticas aqui apresentadas considera-se que o *Google Scholar* constitui uma base de dados apropriada para os investigadores efetuarem buscas nas mais diversas áreas de investigação, uma vez que possui indexada uma grande variedade de meios de comunicação dos resultados. Como instrumento que fornece informação para o cálculo de indicadores bibliométricos possui severas limitações que não o tornam adequado para este fim.

As bases de dados *Sciverse Scopus* e *Thomson Reuters Web of Knowledge* apresentam um conjunto de características que as tornam atrativas no cálculo de indicadores bibliométricos:

- A pesquisa de um determinado conjunto de publicações numa destas bases de dados permite obter informação sobre diversos aspetos. É possível visualizar a lista de autores; o periódico onde os documentos foram publicados com toda a informação relativa ao volume, número, paginação, data de publicação e área científica; o tipo de documentos publicados; o idioma em que se encontram redigidos e o endereço institucional dos autores.
- Um relatório de citações é disponibilizado. Este relatório permite fazer uma análise das citações obtidas pelas publicações ao longo do tempo. O número de citações, o número de auto-citações, o número de artigos citantes, o número de artigos auto-citantes, a média de citações por documento e o índice *h* (Hirsch 2005) são fornecidos por este relatório.
- Um instrumento de análise de dados é disponibilizado permitindo pesquisas mais refinadas sobre os autores, colaborações, grupos de trabalho e áreas de investigação

entre outras. Esta informação refinada poderá permitir a identificação de tendências ocultas e áreas de pesquisa emergentes.

- São também disponibilizados indicadores que permitem avaliar o impacto e o prestígio dos periódicos indexados nestas bases de dados.
- Existe a possibilidade de cada investigador criar o seu próprio perfil na base de dados. Neste perfil o investigador pode incluir todas as suas publicações indexadas, informação sobre a sua afiliação institucional e sobre as áreas de interesse. É ainda possível analisar um conjunto de indicadores que fornecem informação acerca da atividade científica do investigador.

Na secção seguinte é apresentada uma descrição sumária de cada uma das bases de dados.

1.2.1 *Thomson Reuters Web of Knowledge*

Thomson Reuter Web of Knowledge é constituído por cinco bases de dados e um conjunto de instrumentos que permitem obter estatísticas baseadas nas contagens de documentos e citações. Das cinco bases de dados disponíveis apenas será efetuada uma breve descrição do *Web of Science (WoS)* por ter sido a utilizada no presente trabalho. Dos instrumentos que fornecem dados estatísticos será feita uma breve descrição do *Essential Science Indicators (ESI)* e do *Journal Citation Reports (JCR)*.

O *Web of Science (WoS)* permite a pesquisa em mais de 12 000 revistas, mais de 30 000 livros e mais de 148 000 *proceedings* de conferências, simpósios, seminários e *workshops* nas áreas das ciências, ciências sociais, artes e humanidades. Toda esta informação está dividida por sete bases de citações:

- *Science Citation Index Expanded (SCI-EXPANDED)* com documentos indexados entre 1900 e o presente, publicados em mais de 8300 periódicos de 150 áreas.
- *Social Sciences Citation Index (SSCI)* com documentos indexados entre 1956 e o presente, publicados em cerca de 2700 periódicos de 55 áreas;
- *Arts & Humanities Citation Index (A&HCI)* com documentos indexados entre 1975 e o presente, publicados em cerca de 2300 periódicos da área das artes e humanidades;
- *Conference Proceedings Citation Index - Science (CPCI-S)* com documentos indexados entre 1991 e o presente.
- *Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)* com documentos indexados entre 1991 e o presente.
- *Book Citation Index – Science (BKCI-S)* com documentos indexados entre 2005 e o presente.
- *Book Citation Index – Social Sciences & Humanities (BKCI-SSH)* com documentos indexados entre 2005 e o presente.

As bases de citações de *proceedings* possuem indexadas publicações de conferências que tiveram lugar em diversas regiões do mundo. Cerca de 41% destas conferências ocorreram na Europa, 41% ocorreram na América do Norte, 12% na Ásia, 3% na Austrália, 2% na América do Sul e 1% em África.

Existe um grande número de periódicos (maioritariamente revistas) que publicam resultados das atividades de investigação realizadas por todo o mundo. No entanto no WoS não estão indexados todos. Para serem indexados tais periódicos têm que obedecer a um conjunto de critérios (Reuters 2012c):

- Pontualidade nas datas de publicação.
- Devem apresentar títulos informativos, resumos, afiliações dos autores, possuir referências bibliográficas redigidas no alfabeto romano e completas.
- O processo de seleção de documentos deve ser feito utilizando como metodologia a avaliação por pares.
- O conteúdo publicado deve ser algo novo numa determinada área e que não seja já explorado por um dos periódicos indexados.
- A diversidade dos autores e editores também é importante. Periódicos que apresentam diversidade internacional de autores e editores são mais valorizados.
- O número de citações que recebem é também tido em consideração. Em cerca de 80% dos periódicos as auto-citações representam menos de 20% do total de citações. Se o número de auto-citações superar em muito o valor de 20% existe uma avaliação cujo objetivo é identificar se a auto-citação está a ser utilizada para inflacionar as citações alcançadas. Se um comportamento fraudulento é identificado o periódico em questão poderá ser excluído da base de dados.

Esta base de dados tem sofrido várias alterações ao longo dos últimos anos com o objetivo de disponibilizar cada vez mais informação ao utilizador. O número de periódicos indexados tem vindo a aumentar permitindo ao utilizador, nas suas pesquisas, abranger um maior número de estudos relacionados com uma área de interesse. A disponibilização de instrumentos e informação que facilitam os estudos bibliométricos através desta base de dados tem também vindo a ser incrementado.

Journal Citations Reports

O JCR é considerado simultaneamente uma base de dados e um instrumento que fornece análises estatísticas. Nesta base de dados estão indexadas mais de 10 500 revistas que estão classificadas em 232 categorias.

Para cada categoria está disponível uma lista de revistas e um conjunto de indicadores que são baseados nas contagens de citações obtidas pelos documentos publicados. Esta informação está disponível para o período 2000-2011.

Os indicadores disponibilizados são os seguintes:

- Número total de citações.
- *Impact Factor* utilizando uma janela de citação de dois anos. Este indicador fornece, por exemplo, o número médio de citações obtidas em 2011 pelos artigos publicados na revista em 2009 e 2010. O valor obtido será o *Impact Factor* da revista em 2011.
- *Impact Factor* utilizando uma janela de citação de cinco anos. Este indicador fornece, por exemplo, o número médio de citações obtidas em 2011 pelos artigos publicados na revista em 2006, 2007, 2008, 2009 e 2010.

- *Immediacy index* - o número médio de citações obtido no ano de publicação pelos artigos de uma dada revista é o que se pretende com este indicador. O objetivo deste indicador é mostrar a velocidade com que os documentos da revista são citados.
- *Journal Cited Half-Life* - uma revista com *cited half-life* igual a 7 em 2011 significa que os artigos publicados na revista entre 2005-2011 (inclusive) são responsáveis por 50% das citações recebidas pela revista em 2011.
- *Journal Citing Half-Life* - uma revista com *citing half-life* igual a 7 em 2011 significa que 50% dos artigos citados pelos artigos publicados na revista foram publicados entre 2005-2011 (inclusive).
- *Eigenfactor Score* - Este indicador é baseado no número de citações recebidas num ano específico pelos artigos publicados na revista nos cinco anos anteriores, tomando também em consideração as revistas onde os artigos citantes foram publicados. Revistas citantes com maior impacto contribuem de forma mais significativa para determinar o *Eigenfactor Score* do que revistas citantes com menor impacto.
- *Article Influence Score* - com este indicador é possível estudar a influência média dos artigos publicados numa revista ao longo dos últimos cinco anos. O indicador é calculado dividindo o *Eigenfactor Score* pelo número de artigos publicados.

No *JCR* na contagem do número de documentos apenas são contabilizados os documentos classificados como artigos e os artigos de revisão por se considerar que os restantes tipos de documentos são pouco citados. Na contagem de citações todos os tipos de documentos são considerados (Reuters 2012a).

Para além deste conjunto de indicadores é ainda possível obter informação sobre a posição ocupada pela revista na ordenação disponibilizada para cada categoria. A ordenação é baseada no *Impact Factor* para dois anos. A identificação do quartil a que a revista pertence numa dada categoria é igualmente disponibilizado.

Essential Science Indicators

O *ESI* é um instrumento analítico que possui informação sobre as citações e os documentos publicados em vinte e duas áreas de investigação ao longo dos últimos dez anos (Reuters 2012b). Este instrumento permite a obtenção de informação em vários domínios.

Ranking de citações - O *ESI* fornece *rankings* para investigadores, instituições de investigação, países e revistas para cada uma das áreas de investigação. Estes *rankings* estão disponíveis para três indicadores bibliométricos: número de publicações, número de citações e média de citações por documento.

Documentos mais citados - A identificação dos documentos no topo 1% e 0,1% em cada área de investigação é também disponibilizada neste instrumento. No topo 1% estão os documentos publicados e mais citados ao longo dos últimos dez anos. No topo 0,1% estão os documentos publicados nos últimos dois anos, em cada área de investigação, e que obtiveram um número muito elevado de citações nos últimos dois meses com informação disponível acerca das citações. Nesta secção é possível identificar quais os investigadores, instituições, países e revistas que estão nestes percentis.

Análise de citações - Aqui são mostradas várias medidas de citações. Para cada área de investigação estão disponíveis a média de citações por documento ao longo dos últimos dez anos. É determinado o número acumulado de citações desde o ano de publicação até o presente para todos os documentos publicados em um dado ano. O número mínimo de citações que um documento deve acumular para estar presente nos vários percentis em cada área de investigação é igualmente disponibilizado. Para além disto, é ainda possível identificar conjuntos de documentos muito citados num tópico muito especializado.

Neste instrumento não são considerados todos os tipos de documentos. A contagem de documentos engloba os seguintes tipos: artigos, artigos de revisão, *proceedings* e notas de investigação.

O *Thomson Reuters ISI Web of Knowledge* foi durante muitos anos a única base de dados multidisciplinar disponível de indexação de literatura científica. Em 2004 a *Elsevier* lançou o *SciVerse Scopus* e que constitui já uma alternativa.

1.2.2 *SciVerse Scopus*

O *SciVerse Scopus* é uma base de dados multidisciplinar que, para além da indexação de documentos de índole científica, disponibiliza meios que permitem calcular indicadores bibliométricos e realizar análises estatísticas para estudos bibliométricos tal como o *Thomson Reuters ISI Web of Knowledge*. Segue-se uma breve descrição sobre o conteúdo indexado na base de dados, bem como do instrumento que permite obter algumas estatísticas sobre a *performance* dos meios de comunicação indexados.

O *SciVerse Scopus* possui indexado mais de 19 500 periódicos dos quais 18 500 são revistas (1 800 de acesso livre); 400 são *trade publications* e 340 são séries. Este conjunto de fontes de informação resulta em mais de 47 milhões de registos indexados dos quais 4,9 milhões são *proceedings* de conferências. Cerca de 78% dos registos possuem referências até 1996, e os restantes possuem referências até 1823.

Os periódicos indexados no *SciVerse Scopus* pertencem a quatro grandes áreas. Cerca de 33% dos periódicos indexados pertencem à área das Ciências Físicas, 32% às Ciências da Saúde, 30% às Ciências Sociais e 20% às Ciências da Vida. A classificação em cada uma das áreas não é exclusiva.

Ao nível da indexação regional, no *SciVerse Scopus* 54% dos periódicos indexados são europeus, 32% da América do Norte e os restantes 16% pertencem a outras regiões como a América Central e do Sul, África, Austrália entre outras.

A indexação dos diferentes meios de comunicação no *SciVerse Scopus* é efetuada mediante a avaliação de um conjunto de critérios que devem ser respeitados (Scopus 2011, 2013b):

- Devem apresentar diversidade geográfica em relação aos editores e autores dos documentos.
- A revisão por pares deve ser a metodologia aplicada para a seleção dos documentos.

- As referências bibliográficas devem estar redigidas no alfabeto romano.
- O conteúdo deve contribuir de forma significativa para o avanço do conhecimento na área de publicação.
- Os documentos devem permitir uma leitura clara e objetiva por parte do leitor.
- As citações dos periódicos são avaliadas no *SciVerse Scopus*.
- As datas de publicação devem ser respeitadas.
- Os conteúdos devem estar disponíveis *on-line*.

Scopus Journal Analyser

O *Scopus Journal Analyser (SJA)* permite observar a evolução do desempenho dos periódicos indexados no *SciVerse Scopus* ao longo dos anos. O desempenho pode ser medido utilizando os vários indicadores fornecidos por esta ferramenta:

- Número total de citações;
- Número total de documentos;
- Percentagem de documentos não citados;
- Percentagem de artigos de revisão publicados;
- *SCImago Journal Rank (SJR)* – Este indicador, no seu cálculo, tem em consideração o prestígio ou *status*, dos periódicos citantes. Quanto maior o prestígio maior a sua contribuição para a determinação do *SJR*. No cálculo do indicador as citações não são tratadas todas do mesmo modo;
- *Source Normalized Impact per Paper (SNIP)* – No cálculo deste indicador as citações são pesadas de acordo com o fluxo de citações da categoria. A uma citação é atribuído um maior peso quando estamos perante uma categoria onde o fluxo de citações é menor.

Outros aspetos poderiam ser mencionados relativamente a estas bases de dados. No entanto, o objeto de estudo deste trabalho não é a descrição detalhada das bases de dados. Providencia-se apenas uma pequena descrição dos instrumentos disponíveis mais importantes com o objetivo de familiarizar o leitor com as potencialidades do *SciVerse Scopus* e do *Thomson Reuters Web of Knowledge*.

1.2.3 As limitações do *Scopus*¹ e *WoS* na indexação

Se nas ciências consideradas “duras” a análise bibliométrica já é aceite como um instrumento que fornece informação sobre o impacto das atividades de investigação e que permite identificar tendências, por exemplo nas ciências dos computadores, ciências sociais e

¹ A partir deste ponto será apenas utilizado o termo *Scopus* quando se pretende referir a base de dados *SciVerse Scopus*.

humanidades a aceitação deste tipo de análise pelos elementos da comunidade científica é inferior.

Para que uma base de dados seja útil na construção de indicadores bibliométricos é fundamental que disponibilize instrumentos que forneçam informação acerca das citações obtidas por um dado documento e sobre quantidade de documentos indexados. No entanto, por si só esta informação não é suficiente para ser utilizada nas análises bibliométricas. É imprescindível conhecer em que medida as bases de dados representam os documentos publicados ao longo dos anos nas diversas áreas científicas.

As políticas de indexação adotadas pelo *SciVerse Scopus* e *Thomson Reuters Web of Knowledge* determinam algumas diferenças na cobertura dos periódicos de diferentes áreas.

Tal como já foi mencionado anteriormente, os livros são um importante meio de comunicação nas ciências sociais e humanidades. Naturalmente, as referências a livros publicados são mais comuns aqui do que nas restantes ciências (Lariviere, Archambault, Gingras *et al.* 2006). As bases de dados multidisciplinares mais utilizadas indexam essencialmente revistas internacionais (Moed 2005 capítulo 7, Scopus 2013b). Estas características tornam as referidas bases mais adequadas para recolher as citações dos documentos publicados em revistas do que em livros. O estudo publicado por Leydesdorff 2003 mostrou que 79% das citações dos documentos indexados no *Science Citation Index* se referiam a artigos indexados na base de dados. Para os documentos indexados no *Social Science Citation Index* a percentagem determinada foi de 45%.

Por outro lado, as publicações nacionais possuem um papel muito importante nas ciências sociais e humanidades (Hicks 2005), o que torna as listas de referências mais heterogêneas relativamente ao periódico e ao idioma usado. No final tudo isto resulta que a análise de citações poderá não ser a mais adequada para a área das ciências sociais e humanidades quando baseadas nas bases multidisciplinares mais utilizadas.

No entanto, ao facto de as bases de dados mais utilizadas na análise de citações sub-representarem os periódicos com uma orientação nacional não deve ser atribuído um grande negativismo. Se nas ciências sociais um investigador opta por publicar essencialmente em periódicos nacionais como se poderá inferir sobre o verdadeiro impacto das suas atividades de investigação? Mesmo que os resultados publicados apenas tenham um interesse nacional, importará dar a conhecer à generalidade da comunidade científica utilizando meios com uma orientação internacional. Este procedimento permitiria obter críticas de uma audiência de maior dimensão, contribuindo para desenvolver trabalhos mais avançados e mais generalizáveis.

A apresentação dos resultados das atividades de investigação na forma de *proceedings* de conferências tem um papel importante. As conferências oferecem aos investigadores a oportunidade de apresentar e discutir os resultados preliminares da sua investigação com outros investigadores com interesses na mesma área de trabalho. O *feedback* recolhido permitirá melhorar o trabalho apresentado e contribuirá para o desenvolvimento de trabalhos de melhor qualidade no futuro. As divulgações efetuadas nas conferências são normalmente publicadas como resumos ou documentos extensos nos *proceedings* das conferências. Apesar de a utilização dos *proceedings* ter vindo a aumentar ao longo dos últimos anos através da sua inclusão nas bases de dados de referência bibliografia, repositórios e até *website* das conferências, a respetiva disseminação dos resultados das atividades de investigação ainda é

inferior à dos artigos publicados em revistas com avaliação por pares e com uma orientação internacional.

A importância atribuída aos *proceedings* difere entre áreas científicas. Nas ciências dos computadores este tipo de documento representa cerca de 20% das referências mencionadas enquanto que para as ciências sociais e humanidades representa cerca de 5% (Lisee, Lariviere e Archambault 2008). Alguns estudos não recomendam o uso de bases de referência bibliográfica como o *Scopus* e *WoS* na avaliação bibliométrica das ciências dos computadores (Academy 2007; De Sutter e Van den Oord 2012; Meyer, Choppy, Staunstrup *et al.* 2009) por se considerar que estas bases de dados possuem indexados um número de *proceedings* que é muito baixo e portanto não é suficiente para inferir sobre a produção científica nestas áreas.

Perante estes factos, pode-se concluir que a utilização do *Scopus* e *WoS* na análise bibliométrica das ciências dos computadores, ciências sociais e humanidades é limitada.

1.2.4 Recolha da informação no *Scopus* e *WoS*

Inicialmente as bases de dados de referência bibliográfica foram desenvolvidas com o objetivo de fornecer à comunidade científica informação sobre os resultados dos vários estudos efetuados por todo o mundo. Só posteriormente foram introduzidas modificações no sentido de adaptar as bases de dados às análises bibliométricas.

Para que seja possível determinar os indicadores é necessário definir o conjunto de documentos sobre os quais se pretende obter informação. A agregação destes documentos nas bases de dados pode tornar-se uma tarefa complexa. Quando se pretende selecionar o conjunto de documentos indexados para um dado investigador é fundamental ter presente que um indivíduo pode possuir vários nomes científicos (sinónimos) e que vários indivíduos usam o mesmo nome científico (homónimos). O nome científico adotado por um investigador é normalmente composto pelo sobrenome, pela inicial do primeiro nome e em muitos casos os investigadores adotam também as iniciais dos restantes nomes.

O facto de existirem vários nomes científicos associados a um investigador pode dever-se a vários fatores:

- O investigador não é consistente relativamente ao nome científico adotado;
- Existem vários periódicos com políticas bem definidas em relação ao nome científico que o investigador deve apresentar no documento publicado;
- Os investigadores podem alterar o seu sobrenome ao longo da sua carreira, como por exemplo, devido a casamento ou divórcio;
- No caso em que o sobrenome consiste na junção de dois nomes (prática muito comum dos investigadores espanhóis e chineses) os erros na transcrição do nome são mais frequentes;
- A transliteração do nome de investigador para o alfabeto romano poderá incluir alguns erros.

Os dois últimos fatores apontados ocorrem com mais frequência se as referências bibliográficas são introduzidas manualmente nas bases de dados eletrónicas. No caso do

Scopus e do *WoS* para a maioria dos periódicos indexados a informação é fornecida numa forma eletrónica pelas editoras diminuindo a ocorrência destes erros.

Os homónimos são de facto algo de difícil resolução quando se pretende seleccionar o conjunto de documentos indexados de um investigador. A combinação do nome do investigador com outra informação, como por exemplo a instituição de afiliação, poderá ajudar na desambiguação.

Efetuar listas corretas de documentos indexados nas bases de dados de referência bibliográfica para um investigador implica que se detenha algum conhecimento prévio sobre o mesmo. Este conhecimento deve englobar informação sobre o nome científico, sobre as instituições onde desenvolveu e desenvolve atividades de investigação e áreas de trabalho na qual esteve ou está ativo. A obtenção desta informação pode ser feita através da consulta do *Curriculum Vitae* do investigador, quando disponível.

A tarefa de agregação dos documentos poderá revelar-se mais ou menos complexa se associada a instituições de ensino superior ou instituições de investigação. O *Scopus* e o *WoS* apresentam para cada documento um campo com informação sobre a afiliação dos investigadores que aparecem na lista de autores. Estas bases de dados adotaram alguns procedimentos com o objetivo de facilitar a busca dos documentos. Se no documento os endereços indicam alguma hierarquia, no *WoS*, a organização que ocupa o topo da hierarquia é mencionada em primeiro lugar e a restante informação é adicionada de acordo com a posição hierárquica. O *WoS* utiliza ainda uma abreviação e normalização das palavras como se ilustra no exemplo providenciado em seguida.

Exemplo

“Univ Porto, Fac Ciencias, REQUIMTE Dept Quim & Bioquim, P-4169007 Oporto, Portugal.”

Neste caso, uma simples pesquisa por “univ porto” iria permitir a seleção do documento. Para a Universidade Nova de Lisboa podem ser encontrados os seguintes endereços:

- “Univ Nova Lisboa, Fac Ciencias & Tecnol, Dept Fis, P-2825 Monte De Caparica, Portugal.”

- “New Univ Lisbon, Dept Mat Sci, Fac Sci & Technol, P-2829516 Caparica, Portugal.”

Isto evidencia que mesmo com a normalização adotada no *WoS* a pesquisa dos documentos deve considerar as várias possibilidades que existem para a designação da instituição. Por outro lado, os autores nem sempre fornecem informação completa acerca das instituições de afiliação e isto pode levar à não seleção de alguns documentos.

No *Scopus* os endereços são transcritos da publicação original sem nenhum tratamento tornando a recolha da informação ainda mais complexa. A busca de documentos deve também considerar as diferentes designações que podem ser atribuídas a uma instituição.

Dado que cada vez mais é importante contabilizar a produção científica indexada nestas bases de dados, os líderes de instituições de ensino superior e de outras instituições de investigação têm alertado para este facto e sugerido instruções de como os endereços da instituição devem ser mencionados nos documentos.

Após a breve exposição dos cuidados a ter na recolha da informação indexada nestas bases de dados conclui-se que a análise bibliométrica é uma metodologia que requer um trabalho inicial de recolha de dados elaborado e que implica a consideração das várias limitações das bases de referências de bibliografia internacional.

1.3 Relação entre indicadores bibliométricos e a avaliação por pares

A avaliação por pares é um processo que utiliza um conjunto de indivíduos qualificados em uma determinada área científica para avaliar o desempenho. Os painéis de avaliação têm como função formular um conjunto de apreciações relacionadas com a qualidade do objeto submetido para avaliação.

Este tipo de avaliação tem sido aplicado em vários contextos: 1) com elevada frequência pela comunidade académica em avaliações internas; 2) em trabalhos submetidos para publicação por painéis editoriais de periódicos; 3) em projetos de investigação candidatos a financiamento público/privado; 4) em teses de doutoramento.

Nos casos em que este método é empregue para avaliar os trabalhos submetidos para publicação num determinado periódico o objetivo passa pela melhoria da qualidade dos trabalhos através da deteção de pontos fracos e erros. Aqui a avaliação tem em consideração a originalidade e contribuição do trabalho desenvolvido para o avanço do conhecimento na comunidade científica. As avaliações são similares nas situações referidas nos pontos 3 e 4. Nas avaliações internas esta metodologia é utilizada nos concursos cujo objetivo é recrutar indivíduos para os diferentes domínios de trabalho. Nesta situação, são avaliados outros parâmetros para além dos mencionados anteriormente, dependendo do objetivo final. Neste sentido pode afirmar-se que a avaliação por pares é vista como um juízo final sobre aquilo que é confiável.

A procura de uma relação entre os resultados fornecidos pelos indicadores bibliométricos e os resultados da avaliação por pares é algo que tem vindo a ser estudado ao longo dos anos. A título de exemplo apresentam-se em seguida os estudos mais referenciados pela comunidade científica neste tópico.

Os exercícios de avaliação nacional e institucional fornecem resultados de grande valor e que podem ser utilizados para efetuar algum controlo de qualidade. Alguns estudos têm vindo a ser efetuados com base nos resultados do *Research Assessment Exercise (RAE)* no Reino Unido e do *Valutazione Triennale della Ricerca (VTR)* em Itália.

Reale, Barbara e Costantini 2007 exploraram a relação entre os resultados da avaliação por pares no VTR, em três disciplinas (biologia, química e economia), e os resultados providenciados pelos indicadores bibliométricos. O *impact factor* dos periódicos foi o indicador utilizado. Os coeficientes de *Spearman* determinados variam entre 0,44 e 0,48. Abramo, D'Angelo e Caprasecca 2009 também estudaram a relação entre os resultados do VTR obtidos para as ciências duras e dois indicadores bibliométricos, o *impact factor* normalizado e o número total de documentos. Os índices de correlação encontrados neste estudo foram superiores aos obtidos no estudo anterior. Para seis das oito áreas avaliadas todos os índices apresentam valores superiores a 0,6. Para as restantes áreas os valores situam-se entre 0,336 e 0,409. Um estudo semelhante foi efetuado posteriormente por

Franceschet e Costantini 2011, mas com um número superior de indicadores. Os indicadores média de citações por documento e *impact factor* estão correlacionados positivamente e significativamente com as decisões dos painéis de avaliação para a maioria das áreas analisadas (coeficientes de *Spearman* entre 0,32 e 0,81 para a média de citações por documento e coeficientes de *Spearman* entre 0,29 e 0,85 para o *impact factor*). O índice *h* mostrou ser um indicador razoável para discriminar entre documentos classificados num dado nível (excelente, bom, aceitável e limitado) pelo painel.

Aos resultados da avaliação por pares do *RAE* foi também aplicada a análise bibliométrica com o objetivo de identificar se alguma associação poderia ser estabelecida entre os resultados.

Norris e Oppenheim 2003 determinaram o coeficiente de *Spearman* entre o número total de citações e os resultados finais do *RAE* para os departamentos de arqueologia. Os coeficientes encontrados são superiores a 0,7 e significativos. Outros estudos realizados anteriormente para departamentos de outras áreas mostraram resultados semelhantes (Holmes e Oppenheim 2001; Oppenheim 1995, 1997).

Numa avaliação de 56 programas de investigação em física da matéria condensada na Holanda Rinia, van Leeuwen, van Vuren *et al.* 1998 verificaram que o número total de documentos não está relacionado com os resultados da avaliação por pares. Os restantes indicadores (relacionados essencialmente com o impacto) mostraram alguma correlação com as decisões finais dos pares. Os coeficientes de *Spearman* variam entre 0,48 e 0,68.

Aksnes e Taxt 2004 encontraram coeficientes de correlação de *Pearson* entre 0,28 e 0,48 para os resultados da avaliação por pares dos departamentos da Faculdade de Matemática e Ciências Naturais da Universidade de *Bergen* e indicadores bibliométricos de quantidade e qualidade. Neste caso, os indicadores utilizados foram o número de documentos publicados utilizando a contagem fracionada, o número de citações por investigador que compõem o departamento, e indicadores de impacto normalizados baseados nas citações. Das cinco correlações obtidas apenas uma não é significativa ao nível de 5%.

van Raan 2006 usou os resultados da avaliação por pares efetuada para um conjunto de 147 grupos de investigação em química nas universidades holandesas procurando encontrar uma relação com o índice *h* e o *CPP/FCSm* (*crown indicator*) (Moed, Debruin e van Leeuwen 1995). O autor verificou que os indicadores se relacionam, de algum modo, com a avaliação por pares. Os indicadores mostraram ser suficientes para distinguir os grupos classificados com bom ou excelente pelo painel de avaliação do grupo que recebeu a classificação de satisfatório. No entanto, os indicadores não se mostraram tão eficientes na distinção entre os grupos classificados com bom e excelente.

Estudos semelhantes aos apresentados anteriormente mas aplicados ao nível individual não são tão comuns. Isto poderá ser explicado pelos seguintes fatores:

- Para alguns indivíduos o conjunto de documentos publicados poderá ser muito baixo no período em avaliação e a aplicação de análises estatísticas poderá não ser viável pois a violação dos pressupostos das metodologias ocorre com maior frequência;
- Por outro lado, o impacto da produção científica é normalmente efetuado utilizando indicadores baseados na contagem de citações. As distribuições das citações são caracterizadas por uma distribuição muito assimétrica. No entanto, quando os indicadores baseados na contagem de citações são utilizados para descrever o impacto

de grandes universos tais como um país ou uma universidade, a assimetria não afetará tão substancialmente os resultados. Ao contrário, no caso em que os indicadores são aplicados a pequenos conjuntos de documentos poderá ocorrer um viés dos resultados. Isto implica uma atenção especial no cálculo e interpretação dos indicadores a nível individual.

Eugene Garfield em conjunto com outros autores, consideraram um conjunto de estudos (6 estudos) realizados com diferentes amostras e em diferentes períodos com o objetivo de determinar se *rankings* baseados em métricas calculadas para investigadores muito citados poderiam ser utilizados para prever os investigadores que seriam nobelistas. Os autores concluíram que as métricas baseadas nas citações (número total de citações e citações por autor) eram suficientes para distinguir os investigadores que recebiam o prémio daqueles que não o recebiam e permitiam fazer previsões de quais os investigadores que no futuro receberão o prémio (Garfield e Welljamsdorof 1992).

Nederhof e van Raan 1987 estudaram a relação entre a avaliação por pares e os indicadores bibliométricos ao nível individual utilizando um concurso cujo objetivo era a atribuição de um grau de honra a alunos de doutoramento. Os autores encontraram que os candidatos que receberam o título de *cum laude* foram mais produtivos e obtiveram mais citações nos dois anos posteriores à obtenção do grau do que os candidatos que não receberam o grau honroso.

Bornmann e Daniel 2006 estudaram a relação entre a avaliação por pares e a contagem de citações usando os candidatos ao *Boehringer Ingelheim Fonds (BIF)*. Os resultados finais evidenciaram que para os documentos publicados pelos investigadores selecionados é esperado um maior número de citações do que para os documentos publicados pelos candidatos rejeitados.

Bornmann, Wallon e Ledin 2008 consideraram os candidatos ao *Long-Term Fellowship e Young Investigator of the European Molecular Biology Organization* e estudaram a relação entre o desempenho científico dos candidatos e a decisão final dos painéis. Os autores observaram que, em média, para os candidatos selecionados é esperado um maior número de documentos e citações no período subsequente ao ano em que ocorreu a seleção.

Costas, van Leeuwen e Bordons 2010 partindo de um conjunto de nove indicadores que caracterizam o desempenho científico de mais de 1000 investigadores do Consejo Superior de Investigaciones Científicas (CSIC) elaboraram diferentes perfis de investigadores de acordo com o desempenho destes. Os autores construíram três perfis que denominaram de *top class*, *Medium class* e *low class*. Para cada perfil determinaram a percentagem de investigadores que pertencem às classes profissionais (ordem crescente de prestígio) denominadas por *tenured scientists*, *researcher scientists* e *researcher professor*. Os resultados mostram que a categoria profissional predominante no perfil *top class* é a denominada por *researcher professors* e que a mesma é a que menos predomina no perfil *low class*.

Do conjunto de estudos aqui enumerados podem ser retiradas algumas conclusões gerais:

- Correlações positivas e significativas foram encontradas entre os resultados da avaliação por pares e os indicadores bibliométricos. As correlações encontradas nem sempre mostram uma forte associação.

- O uso de indicadores bibliométricos como instrumento auxiliar da avaliação por pares parece justificar-se no entanto mais investigação relativamente a este assunto é necessária.

Nos estudos apresentados a avaliação por pares foi utilizada como método para validação dos indicadores bibliométricos. No entanto, é importante ter presente que a avaliação por pares não é uma metodologia isenta de problemas. Por um lado, o júri utiliza os seus próprios conhecimentos e experiência profissional para proceder à avaliação. As dimensões que devem ser avaliadas são mencionadas, mas a escolha dos instrumentos utilizados para tal efeito ficam ao critério de cada júri. Por outro lado, existe um conjunto de aspetos que podem influenciar esta metodologia no sentido negativo, isto é, introduzindo algum viés.

A presença destas insuficiências foi comprovada em vários estudos. Para evidenciar em que medida podem afetar a avaliação por pares selecionaram-se alguns estudos e apresentam-se os respetivos resultados principais.

A falta de concordância nas avaliações efetuadas por dois avaliadores sobre o mesmo objeto foi observada em várias situações.

Hodgson 1997 usando as propostas submetidas simultaneamente ao *Heart Stroke Foundation (HSF)* e *Medical Research Council of Canada (MRC)* determinou o coeficiente de *Pearson* entre os resultados providenciados pelas duas agências. O coeficiente determinado foi de 0,592 ($p < 0,001$). O autor determinou que para 72,6% das propostas ambas as agências decidiram atribuir financiamento à mesma proposta.

As 3000 propostas submetidas para financiamento ao *ARC* foram estudadas por Jayasinghe, Marsh e Bond 2001, 2003, 2006. Os autores procuraram encontrar uma relação entre a avaliação efetuada por diferentes avaliadores à mesma proposta. Os resultados mostraram baixos níveis de concordância entre os avaliadores, situando-se os valores determinados para o coeficiente de correlação intraclass no intervalo 0,19-0,53.

Reale, Barbara e Costantini 2007 estudaram a concordância entre as avaliações dos avaliadores em quatro disciplinas no *Valutazione Triennale della Ricerca (VTR)*. O coeficiente de *Spearman* foi utilizado para este efeito tendo sido obtidos valores que variam entre 0,25 para a química e 0,46 para a economia.

Estudos semelhantes de concordância foram também efetuados considerando os documentos submetidos para publicação em revistas. Cicchetti 1991 estudou áreas com diferentes características. Por um lado, considerou áreas mais gerais e difusas e, por outro, áreas específicas e bem definidas. As principais conclusões apontam para que nas áreas mais gerais e específicas existe maior concordância no que respeita à rejeição dos documentos do que à aceitação. O comportamento oposto foi observado para as áreas mais específicas e bem definidas.

Considerando duas revistas da área das neurociências clínicas cujos documentos submetidos para publicação foram avaliados por dois avaliadores, Rothwell e Martyn 2000 verificaram que os valores assumidos pelo coeficiente *k* de Cohen eram muito baixos (entre 0,08 e 0,28), refletindo portanto a baixa concordância do processo de avaliação por pares.

O processo de seleção de documentos foi também estudado para a conferência intitulada *UK Academic for Information System (UKAIS)* (Wood, Roberts e Howell 2004). Os autores encontraram também baixos níveis de concordância.

Num estudo efetuado por Bornmann e Daniel 2005 utilizando os resultados da avaliação por pares para a seleção de alunos de doutoramento e pós-doutoramento no *Boehringer Ingelheim Fonds (BIF)* foram verificados valores mais homogêneos. Em 76% dos casos os avaliadores concordaram na seleção ou rejeição do candidato.

O número de propostas avaliadas pode influenciar significativamente os resultados da avaliação por pares. Jayasinghe, Marsh e Bond 2003 mostraram que nas situações em que o júri avalia três ou mais propostas a concordância aumenta. Na realidade, o júri passa a dispor de um maior número de referências que lhe permitem realizar julgamentos mais elaborados baseados na originalidade, qualidade e contribuição das propostas para o avanço do conhecimento.

Até este ponto apenas foram mencionados estudos cujo objetivo foi analisar a concordância da avaliação por pares. No entanto, este não é o único aspeto que influencia a avaliação por pares. As características do(s) investigador(es) ou júri(s), tais como a idade, sexo, instituição de afiliação, posição académica e nacionalidade podem constituir um potencial viés desta metodologia. Os conflitos de interesse poderão também ser considerados um potencial viés. Vários estudos têm sido realizados com o objetivo de determinar se existe uma associação destas características com os resultados da avaliação dos pares. Referem-se, em seguida, alguns desses estudos.

No processo de avaliação realizado pelo *ARC* é permitido aos candidatos a indicação de alguns dos elementos que constituem o júri. Marsh, Bonds e Jayasinghe 2007 estudaram se este procedimento poderia ser um potencial viés do processo de avaliação. A principal conclusão alcançada foi que os elementos do painel mencionados pelos candidatos atribuíam melhores classificações do que os elementos do júri proposto pelo *ARC*.

Num outro estudo, utilizando também os resultados do exercício de avaliação realizado no *ARC*, Marsh, Jayasinghe e Bond 2008 referem que os elementos do júri de nacionalidade Australiana atribuem classificações mais baixas do que os elementos do júri de outras nacionalidades. No entanto, os autores verificaram que os candidatos propunham, essencialmente, elementos do júri de outras nacionalidades, facto que poderia em parte justificar o comportamento descrito anteriormente. Controlando este efeito, os autores verificaram que, de facto, os júris com nacionalidade Australiana atribuem uma cotação inferior.

Os estudos que procuraram uma relação entre a instituição de afiliação e os resultados da avaliação por pares mostram resultados diferentes. Utilizando os dados das avaliações efetuadas pelo *ARC* Marsh, Bonds e Jayasinghe 2007 observaram que as propostas efetuadas pelas universidades mais conceituadas tiveram níveis de aceitação superior. Reale, Barbara e Costantini 2007 utilizando os resultados do *VTR* detetaram não existir qualquer tipo de viés introduzido pelo prestígio da instituição proponente.

Jayasinghe, Marsh e Bond 2003 mostraram que o título académico teve um efeito positivo na atribuição final do resultado na avaliação das propostas submetidas ao *ARC*. Para o caso das ciências ser Professor revelou ter um impacto positivo e significativo na análise final. Por sua vez, nas ciências sociais e humanidades ter o título de Professor não produz um impacto significativo. No entanto, a interação deste aspeto com o estatuto da universidade mostrou ter um impacto significativo. Comportamento oposto foi encontrado nas avaliações do *VTR*. Para

as quatro áreas analisadas (química, biologia, humanidades e economia) foram encontrados valores de prova (*p values*) não significativos ao nível de 5% (Reale, Barbara e Costantini 2007).

Jayasinghe, Marsh e Bond 2003 concluíram que o género não constitui um potencial viés da avaliação por pares. Relativamente à idade os autores verificaram que não existe um impacto significativo na avaliação nas ciências, mas o comportamento oposto foi verificado para as ciências sociais e humanidades. Este estudo foi também baseado nos resultados das avaliações levadas a cabo no *ARC*.

Relativamente aos aspetos que podem introduzir potenciais viés na avaliação por pares os estudos mencionados não permitem uma conclusão geral. O mais longe que se pode ir nas conclusões é que a influência destes aspetos depende de vários fatores tais como a área e cultura científica do sistema que é avaliado.

A revisão da literatura referida nesta secção mostra que a relação entre os resultados da avaliação por pares e os fornecidos pelos indicadores bibliométricos é um assunto que tem merecido atenção por parte da comunidade científica. No entanto, algum cuidado deve ser tomado no momento da interpretação dos resultados pois os indicadores não conseguem traduzir toda a complexidade e multidimensionalidade do desempenho científico.

1.4 O uso de análises bibliométricas em Portugal

A utilização de indicadores bibliométricos para descrever o desempenho do sistema científico de um país é uma prática aceite pela comunidade científica e é já executada em vários países. Em Portugal o sistema científico sofreu um crescimento considerável nos últimos anos, mas a falta de pessoal qualificado na área da bibliometria explica a inexistência de indicadores de desempenho descrevendo a evolução do sistema ao longo dos anos. Os poucos indicadores disponíveis são providenciados pela Direção Geral de Estatísticas da Educação e Ciência (DGEEC). Estes indicadores são construídos com base na informação fornecida pela *Thomson Reuters* (DGEEC 2012) e apenas são apresentados para o nível nacional. Para as instituições de ensino superior apenas são conhecidos os indicadores de desempenho providenciados pelo *SCImago Institutions Ranking (SIR)* e o *Leiden Ranking*.

Em consequência da ausência de informação existem já instituições de ensino superior que requereram a entidades especializadas nesta área relatórios que descrevem detalhadamente o desempenho científico da instituição. A Universidade de Nova de Lisboa requisitou nos últimos 6 anos dois relatórios ao *Centre for Science and Technology Studies* da Universidade de *Leiden* (Visser, van Raan e Nederhof 2007; van Wijk e Visser 2012). Recentemente o mesmo procedimento foi adotado pela Fundação para a Ciência e Tecnologia com o objetivo de obter informação sobre o desempenho das unidades de investigação e desenvolvimento (I&D) e laboratórios associados (LA).

Outras instituições, como a Universidade do Porto, publicam anualmente um relatório elaborado internamente (no caso pelo Serviço de Melhoria Contínua da Reitoria) no qual também se efetua uma descrição detalhada da respetiva produção científica.

Alguns laboratórios associados também utilizam indicadores bibliométricos para analisar o seu desempenho científico e, em alguns casos, para apoiar a distribuição do financiamento

destinado a atividades de investigação. O Instituto de Engenharia de Sistemas e Computadores do Porto (INESC Porto), o Centro de Investigação em Materiais Cerâmicos e Compósitos (CICECO), a Rede de Química e Tecnologia (REQUIMTE) e o Instituto de Biologia Molecular e Celular (IBMC) são exemplos de laboratórios que recorrem a indicadores bibliométricos.

Os estudos baseados em análises bibliométricas em Portugal são poucos e muito recentes o que sugere que muito pode ainda ser feito no âmbito deste domínio.

1.5 Motivação e objetivos da investigação

A avaliação por pares constitui a metodologia mais usada em todo o mundo quando o objetivo é aferir o desempenho do sistema científico de instituições de ensino superior, instituições de investigação e outros. Devido aos elevados custos e tempo requerido por esta metodologia, apenas uma pequena fração do total de documentos disponível é normalmente avaliada. No entanto, a necessidade de avaliações mais frequentes e mais extensas tornam as análises bibliométricas atrativas. A avaliação de todo o conjunto de documentos permitirá uma visão mais ampla do desempenho. Por outro lado as análises bibliométricas permitem atualizações rápidas e têm baixos custos.

Nas universidades portuguesas os concursos públicos para as posições académicas utilizam a avaliação por pares. Nestes concursos vários parâmetros são avaliados, de entre os quais o desempenho científico. Dependendo da área científica e do período decorrido desde o início da carreira, os candidatos a estes lugares podem apresentar, longas listas de publicações, o que torna a tarefa de avaliação do desempenho científico bastante morosa. A disponibilização de um conjunto de indicadores bibliométricos que permitissem caracterizar as várias dimensões do desempenho científico seria útil para ajudar os membros do júri na formulação das decisões finais. De facto, o uso dos indicadores ajudaria a ultrapassar algumas das limitações apresentadas pela avaliação por pares das quais se destacam:

- Nestes concursos ao painel de avaliação é fornecida informação acerca dos parâmetros que devem ser avaliados. No entanto, a cada elemento do painel é dada liberdade para o uso dos seus próprios critérios de avaliação. A falta de critérios uniformes para a realização da avaliação faz com que a esta metodologia esteja associada alguma subjetividade.
- A comunidade científica poderá oferecer alguma resistência a novos paradigmas e portanto candidatos que trabalham em áreas emergentes poderão ser penalizados.
- Só a utilização de um painel especializado na área de trabalho de cada candidato seria capaz de identificar o verdadeiro impacto do trabalho desenvolvido ao longo dos anos. Nestes concursos o painel de avaliação designado possui conhecimento suficiente para identificar se as atividades de investigação do candidato são desenvolvidas numa área com particular interesse para a sociedade, mas poderá existir uma maior distância entre o julgamento efetuado, relativamente ao impacto, por este painel e aquele providenciado por um painel mais especializado. A utilização de indicadores bibliométricos que permitem inferir sobre o impacto poderá auxiliar os painéis de avaliação destes concursos.

As limitações apresentadas pela avaliação por pares e as vantagens que podem advir da utilização de análises bibliométricas constituíram a principal motivação para o estudo que aqui se descreve e que trata da aplicabilidade dos indicadores bibliométricos como instrumento auxiliar da avaliação por pares.

A aplicação de indicadores bibliométricos a indivíduos não é completamente aceite pela comunidade científica. No entanto, são necessários mais estudos para determinar quão longe os indicadores bibliométricos podem ir quando aplicados a indivíduos. A revisão da literatura efetuada aponta para um baixo número de estudos que visam este objetivo. Por outro lado, nestes estudos utilizaram-se um baixo número de indicadores bibliométricos e também métodos estatísticos cujos resultados finais fornecem informação limitada acerca da verdadeira contribuição das análises bibliométricas para a avaliação por pares. Estes aspetos contribuíram para reforçar as motivações iniciais.

Consequentemente e tendo em conta que se pretende contribuir para o avanço do conhecimento da comunidade científica neste domínio, o principal objetivo deste trabalho é o seguinte:

- Modelar a decisão final dos painéis de avaliação dos concursos para Professor Associado e Catedrático realizados ao longo dos últimos anos nas universidades portuguesas utilizando indicadores bibliométricos. A decisão final do painel resulta da avaliação de um conjunto de parâmetros de entre os quais faz parte o desempenho científico. As listas de publicações dos candidatos estão, normalmente, acessíveis através do *Curriculum Vitae*, enquanto que a informação relacionada com os restantes parâmetros avaliados é de obtenção difícil. Utilizando a lista de publicações de cada candidato pretende-se determinar um conjunto de indicadores de desempenho e demonstrar a aplicabilidade das técnicas bibliométricas como instrumento de apoio à decisão na avaliação por pares.

Para atingir o objetivo enumerado foi necessário procurar respostas para um conjunto de questões de investigação.

- 1) De entre as bases de dados multidisciplinares de referências bibliográficas qual a mais adequada para o fim em vista?
- 2) Quais os indicadores que podem ser utilizados na descrição da produção científica de um indivíduo?
- 3) Será possível a definição de um modelo baseado em indicadores bibliométricos que traduza as decisões tomadas pelos diferentes painéis de avaliação?
- 4) O modelo obtido será robusto em relação ao conjunto de concursos utilizados para a parametrização?
- 5) Qual a influência de cada indicador no resultado final providenciado pelo modelo?
- 6) Em que medida o modelo pode prever a decisão final dos painéis de avaliação dos concursos?
- 7) Melhoraria a qualidade das previsões do modelo se apenas a produção científica fosse avaliada nestes concursos?

1.6 Delineamento da tese

Nesta secção resumem-se as etapas percorridas para responder às questões de investigação mencionadas e posterior concretização dos objetivos.

Estudo das bases multidisciplinares

Dada a impossibilidade de realizar o estudo no espaço de tempo disponível recorrendo às duas bases de dados multidisciplinares mais utilizadas para análises bibliométricas foi necessário selecionar uma para a recolha da informação necessária acerca da produção científica a ser utilizada. A produção científica do conjunto de universidades pertencentes ao Concelho de Reitores das Universidades Portuguesas (CRUP) serviu de base para analisar as diferenças que podem ser encontradas em cada uma das bases multidisciplinares *Scopus* e *WoS*. Este estudo permitiu inferir sobre as limitações destas bases de dados e em que medida são complementares uma da outra. Interessará selecionar aquela base de dados que melhor representa a produção científica de uma instituição multidisciplinar. O estudo considerou apenas universidades uma vez que os indicadores bibliométricos foram aplicados a um conjunto de investigadores que desenvolvem as suas atividades neste ambiente académico.

Nesta etapa pretendeu encontrar-se uma resposta para a primeira questão de investigação.

Indicadores bibliométricos

Após a seleção da base de dados procedeu-se à seleção de alguns indicadores de desempenho já existentes. Outros indicadores foram desenvolvidos ou constituem uma melhoria dos já existentes. Nesta etapa descreve-se o processo de normalização utilizado e efetua-se uma comparação dos novos indicadores desenvolvidos com outros já existentes.

Procurou-se, nesta fase, uma resposta para a segunda questão de investigação.

Definição do modelo

Uma vez definidos todos os aspetos necessários para a realização da análise estatística desenvolveu-se o modelo pretendido recorrendo a modelos de escolha discreta (*discrete choice models*). Para este fim utilizou-se uma amostra constituída pelos dados de um conjunto de 27 concursos para Professor Associado e Catedrático que tiveram lugar em algumas universidades portuguesas. No final do estudo foi possível obter três modelos. Para cada um deles foi estudada a sua robustez, a probabilidade de cada um ser o melhor na explicação das decisões finais dos painéis e a influência de cada um dos indicadores no resultado final.

A realização desta etapa permitiu responder às perguntas de investigação 3, 4 e 5.

Poder de previsão do modelo

Após a definição dos modelos estudou-se o respetivo poder de previsão aplicando várias metodologias e comparando os resultados, sempre que possível, com um cenário para o qual não existe informação sobre a produção científica do candidato (cenário em que os candidatos seriam ordenados aleatoriamente).

Os resultados obtidos nesta fase permitiram responder às questões de investigação 6 e 7.

Conclusões gerais

Nesta secção são apresentadas as principais conclusões elaboradas a partir dos resultados obtidos. São apresentadas ainda algumas sugestões para a realização de trabalhos futuros.

Lista de publicações

Neste capítulo é apresentada a produção científica publicada em periódicos com circulação internacional e as comunicações em painel apresentadas em conferências que resultaram do trabalho desenvolvido.

Capítulo 2

Estudo das bases multidisciplinares

No presente capítulo são apresentados os resultados de um estudo de comparação das bases de dados *Scopus* e *WoS*.

Durante muitos anos o *WoS* foi a única base de dados com documentos de índole científica indexados e com as respetivas citações tendo adquirido um papel fundamental como instrumento de análise bibliométrica. No entanto em 2004 o *Scopus* foi lançado no mercado e desde logo se começou a afirmar, constituindo hoje uma alternativa ao *WoS* como instrumento de análise bibliométrica.

Desde o surgimento do *Scopus* que se sentiu necessidade de estudar estas duas bases de dados no sentido de comparar as funcionalidades fornecidas pelas mesmas aos mais diversos utilizadores. Os estudos efetuados podem ser divididos em dois grupos principais. Em seguida são mencionados alguns dos estudos.

A) Estudos relacionados com os periódicos indexados nas duas bases de dados

Estes estudos procuraram diferenças entre os periódicos indexados no *WoS* e no *Scopus*. Gavel e Iselid 2008 comparam os periódicos indexados no *Scopus* com os indexados no *WoS* e verificou que cerca de 6256 periódicos, que representavam 46% do total indexados no *Scopus*, não estavam indexados no *WoS*. Para 16% dos registos indexados no *WoS* os mesmos não foram encontrados no *Scopus*.

A comparação dos periódicos indexados no *Scopus* e no *WoS* da área da oncologia mostrou que todos os periódicos indexados no *WoS* também estavam no *Scopus*. O mesmo não foi verificado para os periódicos indexados no *Scopus*. Os autores verificaram que 46% dos periódicos indexados no *Scopus* não estavam no *WoS* (Lopez-Illescas, de Moya-Anegón e Moed 2008).

Norris e Oppenheim 2007 utilizando o conjunto de publicações submetidas ao 2001 *Research Assessment Exercise*, na área das ciências sociais, verificaram que o *Scopus* tinha indexados 50,6% e o *WoS* 43,4% do total de periódicos onde os documentos submetidos para avaliação haviam sido publicados. Os autores observaram ainda que o *WoS* tinha indexados 107 periódicos que não estavam no *Scopus*. O *Scopus* tinha indexados 500 periódicos que não constavam do *WoS*.

Os resultados dos estudos aqui apresentados como exemplo apontam para a maior dimensão do *Scopus* relativamente ao número de periódicos indexados, no entanto existem periódicos indexados no *WoS* que não se encontram no *Scopus*.

B) Estudos relacionados com a análise das citações recolhidas nas duas bases de dados

Os estudos a este nível pretenderam identificar quais as limitações e potencialidades das bases de dados quando o objetivo é medir o impacto. A lista de periódicos indexados no *Scopus* é bastante superior à lista indexada no *WoS*. No entanto o *Scopus* para uma percentagem elevada dos documentos apenas possui citações disponíveis após o ano 1996.

Um estudo baseado nas publicações de 22 investigadores de topo, da área *human-computer interaction*, do projeto *EQUATOR* permitiu observar que do total de documentos citantes encontrados no *Scopus* e *WoS* 93% estavam indexados no *Scopus* e apenas 54% estavam indexados no *WoS*. A principal explicação apontada pelos autores do estudo para este resultado é o maior número de *proceedings* indexados no *Scopus*. No entanto as

ordenações obtidas usando a contagem de citações fornecidas por cada uma das bases de dados são semelhantes (coeficiente de *Spearman* igual a 0,970). O índice *h* foi também determinado e valores mais elevados foram encontrados no *Scopus* (Meho e Rogers 2008).

As diferenças obtidas nos valores do índice *h* quando calculado utilizando o *Scopus* e *WoS* foram também estudadas por Bar-Ilan 2008 para um conjunto de 40 investigadores israelitas muito citados. Para este conjunto de investigadores verificou-se que as diferenças observadas para os valores do índice *h* não eram significativas.

Para as publicações pertencentes a 25 membros do departamento de *Library and Information Science* da Universidade de Indiana Meho e Yang 2007 recolheram as citações no *WoS* e no *Scopus*. Posteriormente os autores elaboraram um *ranking* baseado na contagem de citações obtidas. Os resultados mostraram uma sobreposição significativa dos indivíduos posicionados nas posições de topo e nas últimas posições. Para as posições medianas verificou-se que existia uma grande divergência. Estudo semelhante foi também efetuado por Bar-Ilan, Levene e Lin 2007 utilizando investigadores israelitas com um elevado índice de citação. Os autores encontraram ordenações muito semelhantes quando a contagem de citações é baseada no *WoS* ou no *Scopus*.

A contagem de citações utilizando as duas bases de dados foi efetuada para os artigos publicados nas áreas de oncologia e física condensada em 1993 e 2003 por Bakkalbasi, Bauer, Glover *et al.* 2006. Os resultados mostraram que para a área da oncologia no *WoS* a média de citações por documento é superior à obtida para o *Scopus* considerando o ano de publicação de 1993. Para o ano 2003 observou-se o comportamento oposto. Na área da física condensada os autores concluíram que nos dois anos analisados a média de citações por documento era superior no *WoS*. O resultado obtido para o ano 1993 já era esperado pois o *Scopus* apenas fornece informação sobre citações a partir do ano 1996.

Os estudos apresentados permitem concluir que de facto as bases de dados apresentam algumas diferenças e que as mesmas não podem ser colocadas de lado quando se pretende fazer análises bibliométricas. Nos estudos mencionados a extensão das listas dos periódicos indexados em cada uma das bases de dados parece ser a principal justificação para as diferenças observadas nas contagens dos documentos e nos valores obtidos para os indicadores de impacto. No entanto considera-se importante estudar em detalhe que outros fatores podem justificar as diferenças observadas nas análises bibliométricas baseadas na informação providenciada por cada uma das bases. Isto foi explorado no presente capítulo utilizando um conjunto de universidades portuguesas para as quais alguns indicadores bibliométricos foram calculados baseados na produção científica indexada no *Scopus* e *WoS*.

2.1 Amostra

Amostra de universidades

No estudo apresentado neste capítulo foram utilizados os documentos publicados, entre 2000-2007 e indexados no *WoS* e *Scopus*, pelas 16 universidades pertencentes ao Conselho de Reitores da Universidades Portuguesas (CRUP). O conjunto de concursos utilizados

pretendiam todos a seleção de um ou mais candidatos para posições académicas em algumas destas universidades daí a utilização deste conjunto de universidades neste estudo.

Amostra de investigadores

Uma amostra constituída por 210 investigadores holandeses, foi utilizada para estudar as implicações da recolha dos documentos indexados apenas em três das bases de citações que integram o *WoS (SCI-EXPANDED, SSCI e A&HCI)*. Foram selecionados investigadores holandeses porque na Holanda os líderes das universidades têm uma prática de recrutamento de investigadores de elevada qualidade científica, procurando responder às rigorosas exigências internacionais nesta matéria. Por outro lado, as políticas de financiamento nacional visam não só satisfazer as necessidades estratégicas, mas também proporcionar o apoio adequado aos investigadores com novas ideias. A Holanda não é o único país com estas características existindo outros, como a Suíça ou Dinamarca, que também poderiam ter sido escolhidos. Em suma, foi selecionado um país onde são recrutados investigadores de elevada qualidade e onde existem as condições favoráveis para o desenvolvimento da carreira. Julga-se que nestas condições é possível obter conclusões mais rigorosas do estudo.

Os investigadores selecionados para esta amostra respeitam o seguinte conjunto de critérios:

- possuem como afiliação uma instituição de ensino superior holandesa;
- são os investigadores holandeses com um maior número de documento nas categorias selecionadas no período 2003-2007. Estes possuem pelo menos 50% dos documentos publicados numa das seguintes categorias: *Biochemistry Molecular Biology; Chemistry, Organic; Chemistry, Physical; Computer Science Theory & Methods; Engineering, Chemical; Engineering, Electrical & Electronic; Environmental, Sciences; Mathematics, Applied; Oncology; Pharmacology & Pharmacy; Physics, Applied; Physics, Condensed Matter;*
- o número de documentos publicados neste período é superior a 10. Este limite foi estabelecido para evitar um número muito baixo de documentos por investigador e portanto com pouco significado aquando da interpretação dos resultados;
- a soma da produção científica destes investigadores representa entre 10% e 16% do total de documentos publicados entre 2003-2007 da categoria onde possuem 50% ou mais da sua produção científica.

2.2 Metodologia

2.2.1 Pesquisa da produção científica no *Scopus* e *WoS* para instituições

Na realização deste estudo utilizou-se a informação sobre a produção científica indexada no *Scopus* e *WoS* de cada uma das universidades do CRUP. Registe-se que de entre o conjunto das universidades pertencentes ao CRUP apenas a Universidade Nova de Lisboa e a

Universidade do Porto disponibilizam no seu portal informação sobre a quantidade e impacto da produção científica indexada no *WoS*.

Assim, perante a ausência de informação sobre a produção científica da maioria das universidades foi necessário desenvolver uma metodologia de pesquisa que, quando aplicada a estas bases de dados, permitisse obter a maior informação possível sobre os documentos indexados e num pequeno espaço de tempo. Na Figura 1 apresentam-se esquematicamente as diversas etapas associadas à metodologia de pesquisa.

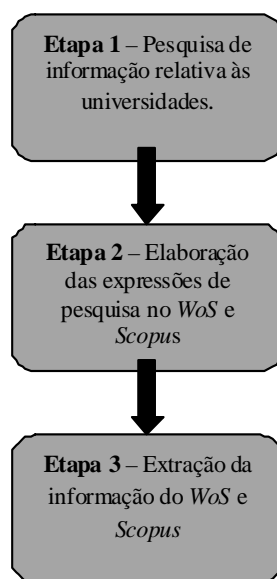


Figura 1. Metodologia de pesquisa utilizada na obtenção da produção científica indexada no *Scopus* e *WoS* das universidades do CRUP.

O método de pesquisa da produção científica das universidades baseia-se na informação contida nos endereços indicados pelos autores dos documentos publicados por cada universidade. A cada instituição de ensino superior associam-se todos os documentos que no endereço de um dos autores contém uma referência clara à universidade, a uma das suas unidades orgânicas ou dos seus grupos de investigação. Com esta metodologia é possível obter informação sobre a produtividade de uma universidade respeitando integralmente os endereços apresentados em cada documento.

Uma descrição detalhada de cada uma das etapas necessárias para obtenção desta informação é apresentada em seguida.

Etapa 1 - Pesquisa de informação relativa às universidades

Uma vez que a metodologia de pesquisa se baseia nos endereços indicados pelos autores é necessário possuir informação sobre quais as unidades orgânicas e grupos de investigação associados a cada universidade. Esta informação foi obtida através de pesquisas no portal de cada universidade.

Etapas 2 – Elaboração das expressões de pesquisa no WoS e Scopus

A recolha dos documentos indexados nestas bases de dados para cada universidade foi efetuada utilizando expressões (ou chaves) de pesquisa que são construídas com base nas diferentes designações que a universidade, as unidades orgânicas e instituições de investigação associadas podem assumir. O *Scopus* e o *WoS* disponibilizam um conjunto de instrumentos que permitem a elaboração destas expressões. No exemplo indicado em seguida é apresentada a expressão de pesquisa utilizada no *WoS* para a Universidade Nova de Lisboa. No *Scopus* utiliza-se o mesmo tipo de expressões.

Exemplo

py=2000-2007 and **cu**=portugal and **ad**=(**"univ nova lisboa"** or **"new univ lisbon"** or **"nova univ"** or **"new univ"** or **unl** or **(fct same caparica)** or **cefitec** or **cenimat** or **itqb** or **"inst tec* quim* biol*"** or **ihmt** or **"hig* med* trop*"** or **"doen* trop*"** or **cepr** or **"univ nouv* lisb*"** or **fcml** or **(**"inst tec* quim*"** same oeiras))**

Esta expressão permite a pesquisa de todos os documentos indexados no *WoS* publicados entre 2000 e 2007. Sempre que uma das designações consideradas é mencionada no campo de endereços (**ad**) o registo é selecionado. Se um documento publicado no período considerado apenas menciona Portugal no endereço e a sigla FCT (Faculdade de Ciência e Tecnologia) em simultâneo com a sua localização (Caparica) será também considerado como pertencente à Universidade Nova de Lisboa. No campo destinado ao endereço o nome da universidade pode não ser mencionado, mas o conhecimento da sua estrutura permite obter documentos que foram publicados por indivíduos que de algum modo pertencem à universidade. Se à designação “FCT” não fosse associada a sua localização iriam ser selecionados não só todos os documentos da Universidade Nova de Lisboa mas também os que no campo do endereço não mencionam a Universidade de Coimbra: é que ambas as universidades (Universidade Nova de Lisboa e Universidade de Coimbra) possuem uma unidade orgânica designada por Faculdade de Ciência e Tecnologia (FCT).

A construção das expressões de pesquisa torna-se ainda mais complexa quando se consideram as universidades que oferecem cursos de medicina. A relação que existe entre tais universidades e os hospitais que apoiam as unidades orgânicas de medicina coloca problemas de análise por não ser clara a fronteira entre os dois domínios.

As três faculdades de medicina portuguesas mais antigas tiveram “hospitais universitários” com dupla tutela que passaram para o Ministério da Saúde há alguns decénios. Contudo, estes hospitais continuam a desempenhar funções de apoio à formação médica. Para o conjunto das universidades aqui consideradas a definição do domínio é mais complexa para a Universidade de Coimbra. Em Coimbra os hospitais mantiveram a denominação de “Hospitais da Universidade de Coimbra” apesar de terem um estatuto de tutela análogo ao dos antigos hospitais universitários de Lisboa (Hospital de Santa Maria) e do Porto (Hospital São João).

Para a Universidade de Coimbra todos os documentos que mencionam “Hospitais da Universidade de Coimbra” são contabilizados como pertencentes à universidade. Para a Universidade do Porto apenas são considerados os documentos que mencionam o Hospital de

São João em simultâneo com a Universidade do Porto ou alguma das suas unidades orgânicas e instituições de investigação. O mesmo procedimento foi adotado para a Universidade de Lisboa. Perante a impossibilidade de aplicar um procedimento semelhante à Universidade de Coimbra é inevitável uma sobreavaliação da produção desta universidade.

Com esta metodologia foram desenvolvidas 32 expressões de pesquisa, duas expressões para cada universidade, que permitem contabilizar a produção científica indexada no *Scopus* e *WoS*.

Etapas 3 – Extração da informação do WoS e Scopus

Na pesquisa foram considerados todos os tipos de documentos e todos os idiomas em que se encontram redigidos. Para o caso do *WoS* os documentos estão indexados nas bases de citações *SCI-EXPANDED*; *SSCI* e *A&HCI*.

Após a seleção os documentos foram extraídos das bases de dados para posterior tratamento. Foi recolhida informação sobre os autores, título, periódico, endereço dos autores, tipo de documento, ano de publicação, número de citações obtidas e categorias em que os periódicos estão classificados.

2.2.2 Pesquisa da produção científica no WoS para investigadores

Após a identificação da base de dados mais adequada, que como se verá mais à frente é o *WoS*, tornou-se necessário fazer um estudo relativamente às diferentes bases de citações que compõem o *WoS*. Neste estudo utilizou-se a produção científica indexada no *WoS* de cada um dos investigadores holandeses (amostra utilizada). A recolha desta informação apenas foi possível com o desenvolvimento de expressões de pesquisa semelhantes às desenvolvidas para as universidades. Nestas expressões foi necessário introduzir informação sobre os investigadores em relação a vários aspetos. No esquema apresentado na Figura 2 ilustra-se a metodologia utilizada para a pesquisa referente a indivíduos no *WoS*.

O método de pesquisa da produção científica dos investigadores baseia-se, essencialmente, na informação contida no campo destinado ao endereço do autor e no campo destinado ao nome. A cada investigador associam-se todos os documentos que contêm no campo dos autores uma referência clara ao investigador e no campo de endereços uma referência à afiliação do investigador no período analisado. Com esta metodologia é possível obter informação sobre a produtividade de um investigador ao longo do tempo. Verificou-se que, em alguns casos esta informação não era suficiente para fazer a identificação de todos os documentos indexados. Nestas situações foi necessário complementar as expressões de pesquisa utilizando a informação contida em outros campos.

Uma descrição detalhada das etapas necessárias para a extração da informação é apresentada em seguida.

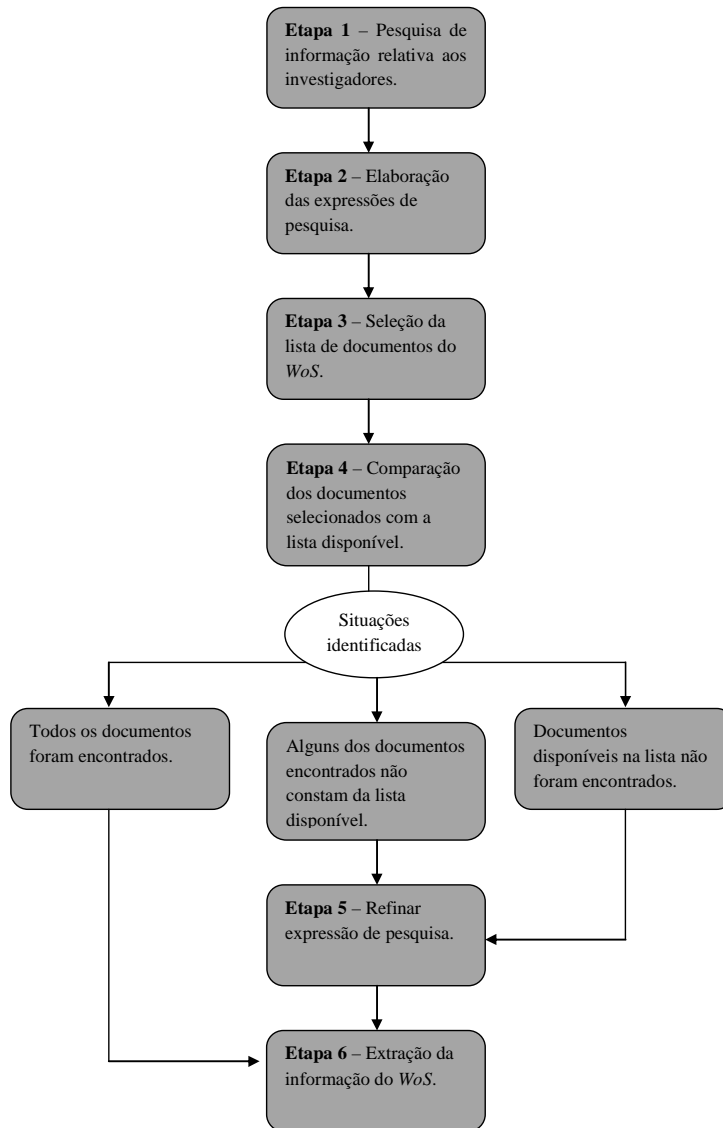


Figura 2. Metodologia de pesquisa utilizada na obtenção da produção científica indexada no WoS dos investigadores.

Etapa 1 – Pesquisa de informação relativa aos investigadores

Para desenvolver as expressões de pesquisa foi necessário recolher um conjunto de informação relativa a vários aspetos complementares. Esta recolha foi efetuada utilizando os currículos dos investigadores quando disponíveis. Nas restantes situações uma pesquisa no portal das instituições de afiliação e nas páginas pessoais dos investigadores permitiu obter a informação necessária.

A informação recolhida relativamente a cada investigador inclui os pontos seguintes:

- a(s) afiliação(ões) de cada investigador ao longo dos anos;
- o(s) nome(s) científico(s) utilizado por cada investigador para se identificar como autor de um dado documento;
- lista de publicações para cada investigador;

Nesta etapa procurou-se, ainda, por outros investigadores que pertencem à(s) mesma(s) instituição(ões) do investigador e que possuem nomes científicos semelhantes.

A informação mencionada foi utilizada posteriormente na elaboração das expressões de pesquisa de acordo com a metodologia que se descreve em seguida.

Etapa 2 – Elaboração das expressões de pesquisa

A informação recolhida na etapa 1 permitiu a construção de expressões de pesquisa como aquela que se mostra no exemplo seguinte.

Exemplo

au=Martins BAB and **ad**=(**"univ porto"** or ((**ciq** or **"ctr invest quim"** or **"chem invest ctr"**) same (oportor or portor))) and **py**=2007

Esta expressão de pesquisa permite obter todos os documentos publicados pelo autor “Martins BAB” (nome fictício) indexados no WoS, publicados no ano 2007, e que possuem uma referência no campo de endereços (**ad**) à Universidade do Porto, ou ao Centro de Investigação em Química da Universidade do Porto.

Na amostra de investigadores utilizada são poucos os casos em que na afiliação não é mencionada a universidade. Perante isto, no campo de endereços é colocado inicialmente as possíveis referências à universidade. Quando a universidade não é mencionada a expressão de pesquisa é refinada para incluir a unidade orgânica, departamento ou centro de investigação a que o investigador está associado. Este procedimento também é adotado no caso em que dois investigadores pertencentes à mesma universidade utilizam nomes científicos iguais ou muito semelhantes. Para as situações em que ambos os investigadores pertencem à mesma unidade orgânica, departamento e centro de investigação a recolha de informação tem de ser efetuada documento a documento na base de dados.

Etapa3 –Seleção da lista de documentos do WoS

A seleção dos documentos no WoS foi efetuada utilizando na opção “pesquisa avançada” as expressões de pesquisa desenvolvidas. Todos os tipos de documentos e idiomas foram considerados. Os documentos selecionados estão indexados nas bases de citações *SCI-EXPANDED*, *SSCI*, *A&HCI*, *CPCI-S* e *CPCI-SSH*.

Etapa 4 – Comparação dos documentos selecionados com a lista disponível

Após a seleção, os documentos foram comparados com a lista disponível. A lista disponível refere-se ao conjunto de publicações de um investigador listadas no seu CV, na página pessoal do investigador ou no portal da instituição de afiliação (informação recolhida na etapa 1). Várias situações foram encontradas:

- a) todos os documentos que constituíam a lista disponível foram encontrados no WoS;

- b) alguns dos documentos encontrados não constavam da lista disponível;
- c) documentos disponíveis na lista não foram selecionados com a expressão de pesquisa utilizada.

Por não constarem da lista, os documentos encontrados no ponto b) suscitaram algumas dúvidas. Neste caso observaram-se duas situações:

- 1) os documentos de facto não pertencem ao investigador. O nome científico utilizado pelo investigador é também usado por um outro investigador que pertence a uma instituição diferente, mas que colaborou com um investigador que pertence à mesma instituição do investigador em causa;
- 2) os documentos pertencem na realidade ao investigador, mas não constam da lista disponível.

O ponto c) é justificado pelas seguintes situações:

- 1) o uso de um nome científico ou afiliação diferente. Na pesquisa efetuada acerca do investigador uma ou mais versões do seu nome científico não foi detetada, principalmente se este a usou num número muito pequeno de documentos. Relativamente aos endereços o mesmo foi verificado. Erros na transferência da informação para a base de dados também justificam alguns dos documentos em falta;
- 2) os documentos não apresentavam informação no campo de endereços dos autores.

Etapa 5 - Refinar a expressão de pesquisa

Após identificação dos documentos que constavam da lista mas que não foram selecionados pela expressão de pesquisa procedeu-se a um ajustamento das expressões. De acordo com a situação foram adotados os seguintes procedimentos:

- foram incluídas na expressão uma vertente do nome científico ou da afiliação do investigador que não foi considerada;
- quando o campo de endereços não possui informação efetuou-se uma interseção do nome do investigador com o título do respetivo documento.

Para os documentos selecionados pela expressão de pesquisa e que não constavam da lista de publicações um dos seguintes procedimento foi adotado:

- quando existe um investigador que utiliza o mesmo nome científico do investigador em causa e que colabora com a sua instituição de afiliação o ajustamento da expressão de pesquisa torna-se mais difícil. Para o ano 2007 e anos posteriores o WoS associa o endereço a cada autor do documento como se pode ver na Figura 3. Nesta situação facilmente se identifica se o documento pertence ou não ao investigador em causa;

- para anos anteriores esta associação é inexistente e portanto a desambiguação é ainda mais complexa. Nesta situação os documentos duvidosos são comparados com os documentos já identificados como pertencentes ao investigador. Se nestes documentos existem co-autores em comum com os documentos já identificados então estes são considerados como da autoria do investigador. Para o conjunto de documentos em que não foram encontrados co-autores em comum, são utilizadas as categorias dos periódicos em que os documentos foram publicados. Se pertencerem a categorias dos documentos já identificados, os documentos são considerados da autoria do investigador. Se o investigador desenvolveu atividades científicas durante um pequeno período de tempo em uma outra área este processo pode levar à não consideração de documentos que, de facto, pertencem ao investigador.

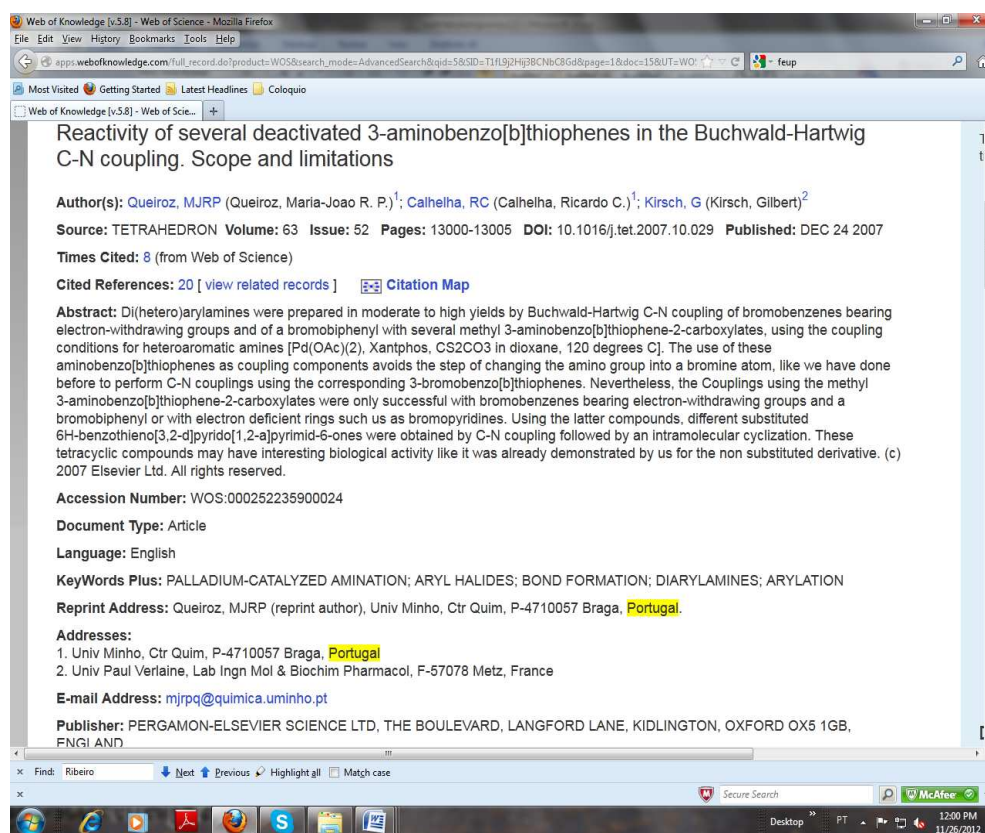


Figura 3. A afiliação de cada autor do documento e restantes aspetos que caracterizam uma publicação indexada no WoS.

Com esta metodologia foi possível selecionar os documentos de cada um dos investigadores indexados no WoS.

Etapa 6 – Extração da informação do WoS

Após o refinamento das expressões de pesquisa quando necessário, procedeu-se à extração dos documentos do WoS para posterior tratamento. A informação extraída relaciona-se com os autores, título, periódico, endereço dos autores, tipo de documento, ano de publicação, número de citações obtidas e categorias em que os periódicos estão classificados.

Com informação sobre a produção científica das universidades e investigadores indexada nas bases de dados dois tipos de indicadores (indicadores de quantidade e impacto) foram determinados na tentativa de encontrar respostas para o pretendido.

2.2.3 Quantificação da produção científica

Para cada universidade foram efetuadas contagens relativas ao número de documentos indexados no período entre 2000 e 2007 no *Scopus* e *WoS*. O período adotado é justificado por duas razões: (1) a produção científica portuguesa cresceu significativamente nos últimos anos e considerar um período de tempo mais longo implica trabalhar com um conjunto de dados muito heterogéneo; (2) as bases de dados sofrem alterações constantes no que diz respeito ao número de periódicos indexados e portanto estudos que implicam tomada de decisões devem ser baseados em dados recentes.

Duas universidades foram selecionadas posteriormente para realizar um estudo detalhado dos documentos indexados em cada uma das bases de dados na tentativa de quantificar as diferenças observadas nas contagens. Os documentos publicados em 2006 pela Universidade de Coimbra e Universidade de Lisboa foram utilizados neste caso. Para estas universidades foi contabilizado o seguinte:

- número de documentos indexados em cada uma das bases de dados;
- número de documentos indexados em simultâneo nas duas bases de dados;
- número de documentos indexados em apenas uma das bases de dados;

Após a determinação dos indicadores mencionados anteriormente procedeu-se a uma inspeção individual dos documentos apenas indexados em uma das bases de dados e identificados os fatores que justificam as diferenças nas contagens.

2.2.4 Quantificação do impacto da produção científica

O impacto associado aos documentos foi também analisado contabilizando as citações obtidas entre 2006 e 2008 (inclusive) dos documentos publicados em 2006 pelas duas universidades. O número médio de citações por documento foi determinado para os seguintes conjuntos:

- documentos indexados em cada uma das bases de dados;
- documentos indexados em simultâneo nas duas bases de dados;
- documentos indexados em apenas uma das bases de dados;

2.2.5 Análise das bases de citações do *WoS*

O *WoS* que inicialmente era constituído por apenas três bases de citações (*SCI-EXPANDED*, *SSCI*, e *A&HCI*) passou a integrar em 2008 mais duas bases de dados de citações, o *CPCI-S* e o *CPCI-SSH*.

A produção científica indexada no *WoS* foi utilizada no cálculo de um conjunto de indicadores bibliométricos para a amostra de investigadores que foram candidatos aos concursos. Esta amostra será descrita no Capítulo 4. O cálculo destes indicadores exige uma elevada quantidade de dados para o processo de normalização como se verá no Capítulo 3. A recolha das citações para os documentos utilizados como referência no processo de normalização é feita utilizando o relatório de citações providenciado no *WoS*. Este relatório apenas permitia a consulta de citações para conjuntos de 5000 documentos no máximo. Desde 2012, este valor passou a ser de 10000. O valor máximo, 5000, era ultrapassado frequentemente com as expressões desenvolvidas se a recolha da informação fosse efetuada nas cinco bases de dados. A opção passou pela utilização de apenas as três bases de citações disponíveis até 2008. O uso de apenas três das bases de citações tem implicações em duas dimensões: na dimensão quantitativa pois estamos a considerar um menor número de documentos; na dimensão impacto uma vez que as citações passam a ser contabilizadas para um menor número de documentos. Com o objetivo de determinar quais as implicações do procedimento adotado os seguintes indicadores foram determinados para a amostra de 210 investigadores holandeses:

- número total de documentos indexados para cada categoria nas bases de citações *SCI-EXPANDED*, *SSCI* e *A&HCI* no período 2003-2007;
- número total de documentos indexados para cada categoria nas cinco bases de citações (*SCI-EXPANDED*, *SSCI*, *A&HCI*, *CPCI-S* e *CPCI-SSH*) no período 2003-2007;
- número de total de citações obtidas entre 2003-2007 pelos documentos indexados nas bases *SCI-EXPANDED*, *SSCI* e *A&HCI* e publicados no período 2003-2007;
- número total de citações obtidas entre 2003-2007 pelos documentos indexados nas bases *SCI-EXPANDED*, *SSCI*, *A&HCI*, *CPCI-S* e *CPCI-SSH* e publicados no período 2003-2007;
- número médio de citações por documento em cada categoria para os documentos indexados nas bases *SCI-EXPANDED*, *SSCI* e *A&HCI* no período 2003-2007;
- número médio de citações por documento em cada categoria para os documentos indexados em *SCI-EXPANDED*, *SSCI*, *A&HCI*, *CPCI-S* e *CPCI-SSH* no período 2003-2007.

Optou-se por indicadores simples uma vez que o objetivo foi identificar como é que categorias com diferentes hábitos de publicação e citação estariam representadas nas diferentes bases de citações.

2.3 Documentos indexados no *WoS* e/ou *Scopus* para uma instituição

A análise aqui efetuada permitiu identificar os vários fatores que para além do diferente número de periódicos indexados no *Scopus* e *WoS* podem justificar as diferenças obtidas nas análises bibliométricas quando se usa as duas bases de dados. A magnitude destas diferenças permitiu identificar em que medida estas bases de dados são complementares uma da outra do ponto de vista de uma instituição multidisciplinar como é o caso das universidades e se o ganho que advém da utilização das duas bases na descrição do desempenho científico dos investigadores de uma instituição justifica o uso de ambas as bases de dados.

O número de documentos indexados em cada uma das bases de dados para as universidades pertencentes ao CRUP é apresentado nas Tabelas 1 e 2 para vários anos.

Tabela 1. Número de documentos indexados no *WoS* e *Scopus* para cada uma das universidades do CRUP².

Instituição	2000		2001		2002		2003	
	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>
ISCTE	5	3	5	5	7	7	18	17
Univ. Aberta	6	7	12	14	7	7	16	24
Univ. Açores	36	37	45	36	32	24	51	58
Univ. Algarve	113	107	135	118	156	134	203	201
Univ. Aveiro	279	282	337	353	427	405	513	575
Univ. Beira Interior	36	43	55	59	55	59	56	64
Univ. Católica	51	45	54	50	57	43	67	70
Univ. Coimbra	458	475	515	473	612	532	678	699
Univ. Évora	53	51	48	39	62	66	85	89
Univ. Lisboa	552	446	534	453	609	518	632	632
Univ. Madeira	18	20	28	26	26	24	36	41
Univ. Minho	192	189	249	229	269	259	339	380
Univ. Nova de Lisboa	371	314	411	309	437	380	460	483
Univ. Porto	684	638	713	637	837	687	901	914
Univ. Téc. Lisboa	747	785	830	776	918	907	986	1047
UTAD	35	33	66	61	70	60	83	99
Portugal	3792	3864	4115	4082	4748	4526	5221	5637

²As pesquisas foram efetuadas em 2008.

Tabela 2. Número de documentos indexados no *WoS* e *Scopus* para cada uma das universidades do CRUP (continuação).

Instituição	2004		2005		2006		2007	
	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>
ISCTE	22	37	18	34	27	41	22	43
Univ. Aberta	15	18	13	19	11	15	7	16
Univ. Açores	48	50	56	60	88	92	63	78
Univ. Algarve	187	203	254	248	293	303	281	292
Univ. Aveiro	644	709	703	838	845	942	697	899
Univ. Beira Interior	73	92	82	101	87	114	92	116
Univ. Católica	69	71	72	69	88	84	98	99
Univ. Coimbra	756	796	741	816	943	969	912	997
Univ. Évora	88	105	109	112	161	164	143	159
Univ. Lisboa	759	765	748	800	1022	1010	872	980
Univ. Madeira	33	32	47	49	59	68	65	74
Univ. Minho	450	495	435	484	581	644	481	658
Univ. Nova de Lisboa	528	531	586	597	717	732	595	672
Univ. Porto	1122	1146	1260	1300	1484	1529	1409	1572
Univ. Téc. Lisboa	1112	1268	1154	1325	1355	1479	1158	1465
UTAD	128	131	134	158	174	177	185	197
Portugal	5991	6372	6407	7038	7990	8378	7983	8254

O *Scopus* possui indexado um maior número de documentos publicados por estas universidades do que o *WoS*. O valor obtido é superior ao obtido no *WoS*, em média, 10% aumentando para 17% se apenas os últimos três anos forem contabilizados.

Neste conjunto de universidades a Universidade Aberta e o ISCTE podem ser consideradas casos atípicos. Para estas universidades a diferença no número de documentos entre as contagens no *Scopus* e *WoS* representam cerca de 34% e 28% do total de documentos indexados no *Scopus*. As universidades em causa são muito especializadas e voltadas para as ciências sociais. Uma vez que a cobertura do *Scopus* relativamente às ciências sociais é superior à do *WoS* (Norris e Oppenheim 2007) o comportamento observado para estas duas universidades é aceitável.

Com o objetivo de entender em maior detalhe o que se pode obter destas duas bases de dados, duas universidades foram analisadas pormenorizadamente. As universidades selecionadas foram a Universidade de Coimbra e a Universidade de Lisboa por serem universidades de dimensão considerável, que possuem as áreas académicas tradicionais, incluindo a medicina.

A contagem detalhada do número e documentos indexados em cada uma das bases de dados, para as universidades escolhidas, no ano 2006 é representada na Tabela 3.

O número total de documentos referenciados em qualquer uma das bases de dados é de 1155 para a Universidade de Coimbra e de 1219 para a Universidade de Lisboa. Dos 1155 documentos encontrados para a Universidade de Coimbra, 84% estão indexados no *Scopus*. No *WoS* foram encontrados indexados 82% do total de documentos. Para a Universidade de Lisboa verificou-se que 83% dos documentos estavam indexados no *Scopus* e 84% no *WoS*.

Os valores obtidos permitem concluir que a proporção do total de documentos indexados em cada uma das bases de dados é muito semelhante. No entanto é importante estudar como estão distribuídos estes documentos. Os dados da Tabela 3 mostram que cerca de 66% dos documentos estão indexados simultaneamente no *Scopus* e *WoS*. A percentagem de documentos que estão apenas indexados em uma das bases de dados varia entre 16% e 18%.

Tabela 3. Número de documentos indexados no *Scopus* e *WoS* para a Universidade de Coimbra e Universidade de Lisboa em 2006.

Documentos Indexados	Universidade	
	Universidade de Coimbra	Universidade de Lisboa
Apenas no <i>Scopus</i>	212	199
Apenas no <i>WoS</i>	186	207
Simultaneamente em <i>Scopus</i> e <i>WoS</i>	757	813
Total	1155	1219

As diferenças obtidas entre as duas bases de dados podem ser consequência dos seguintes aspetos:

- A) as políticas de indexação são próprias de cada base de dados contribuindo isto para que as contagens encontradas sejam diferentes entre as bases;
- B) ocorrência de erros na transferência da informação dos periódicos para as bases de dados.

Uma análise detalhada dos documentos pertencentes a estas universidades foi efetuada na tentativa de verificar em que medida os resultados fornecidos pelas bases de dados são influenciados pelos aspetos mencionados acima. Os resultados obtidos são apresentados na Tabela 4. Na primeira coluna desta tabela são apresentadas as diversas situações que explicam as diferenças obtidas nas contagens dos documentos indexados nestas bases de dados para cada uma das universidades.

Analisando os resultados apresentados constata-se que os aspetos referidos contribuem para as diferenças encontradas nas contagens dos documentos. Para cada um destes aspetos foram observadas várias situações:

A1) Alguns documentos não foram identificados porque o periódico em que foram publicados não se encontrava, à data de pesquisa, indexada no *WoS* ou no *Scopus*. Cerca de 12% e 10% dos documentos da Universidade de Coimbra e Universidade de Lisboa, respetivamente, estavam indexados no *Scopus* e não no *WoS*. Para a Universidade de Coimbra e Universidade de Lisboa 0,5% e 0,2% dos documentos estavam indexados no *WoS* e não no *Scopus*.

A2) Vários documentos, indexados no *Scopus*, não foram encontrados no *WoS* mas sim no *ISI Proceedings*. À data de realização deste estudo o *ISI Proceedings* não estava integrado no conjunto de bases de dados que compunham o *WoS* funcionando como uma base de dados independente. No entanto o *WoS* foi sofrendo alterações e neste momento as bases de dados que constituíam o *ISI Proceedings* estão integradas no *WoS*. A percentagem de documentos

encontrados nesta situação foi de 5% e 1,6% para a Universidade de Coimbra e Universidade de Lisboa respetivamente.

A3) No *WoS* cada estudo apresentado numa conferência é indexado como um documento autónomo, incluindo *meetings abstracts*. No *Scopus* isto não se verifica pois é apresentada a referência à conferência como uma entrada única na base de dados e os trabalhos nela apresentados não são tratados como documentos autónomos. Nesta caso é necessário descarregar e abrir o ficheiro associado à conferência para se obter a informação referente aos estudos apresentados. Nesta situação foram encontrados 7,5% e 7,1% dos documentos indexados para a Universidade de Coimbra e Universidade de Lisboa respetivamente.

B1) Vários erros associados à transcrição dos endereços dos autores e da data de publicação foram observados. Para 1,8% e 3,4% dos documentos indexados para a Universidade de Coimbra e Universidade de Lisboa o *Scopus* possui endereços e data de publicação com erros não permitindo portanto a extração destes documentos pela metodologia de pesquisa aqui definida. Estes erros também foram encontrados no *WoS* em 1,2% e 4,1% dos documentos pertencentes à Universidade de Coimbra e Universidade de Lisboa, respetivamente. Os erros encontrados relativamente aos endereços devem-se à referência incorreta e em alguns casos à não referência da universidade em questão ou do país de origem. Para o caso da Universidade de Lisboa o erro mais comum é a referência no endereço do autor à Universidade Nova de Lisboa e não à Universidade de Lisboa como consta do documento original na base de dados *WoS*.

B2) Diversos documentos não estavam indexados apesar de o periódico (volume, número e suplemento) onde estes foram publicados estar indexado na base de dados. Para a Universidade de Coimbra e Universidade de Lisboa foram encontrados 0,26% e 0,41% dos documentos, respetivamente, indexados no *Scopus* e não no *WoS*. O mesmo comportamento foi observado para 2,2% e 2,7% dos documentos publicados pela Universidade de Coimbra e Universidade de Lisboa que estavam indexados no *WoS* e não no *Scopus*, respetivamente.

B3) Para alguns dos periódicos verificou-se que não estavam indexados todos os volumes, número ou suplementos publicados num dado ano na base de dados *Scopus*. Nesta situação foram encontrados 4,1% e 3,5% do total de documentos da Universidade de Coimbra e Universidade de Lisboa respetivamente.

B4) A duplicação de registos foi também encontrada. Para a Universidade de Lisboa observou-se que dois documentos estavam duplicados no *Scopus*. Estes documentos possuem a mesma informação relativamente a título do documento, autores e periódico, mas existem erros de paginação.

Os resultados sugerem que as políticas de indexação seguidas por cada uma das bases de dados são aquelas que mais contribuem para as diferenças encontradas. Do total de documentos indexados em apenas uma das bases de dados, 65% resultam deste fator.

A perda de informação sobre os documentos publicados por cada uma das universidades é maior no *WoS* se considerarmos o fator políticas de indexação. Por outro lado esta perda é superior no *Scopus* se considerarmos o fator erros na transferência da informação.

Tabela 4. Justificação para a ausência de alguns documentos em uma ou outra base de dados em 2006.

Observações	Universidade de Coimbra		Universidade de Lisboa	
	Nº documentos ausentes no <i>WoS</i>	Nº documentos ausentes no <i>Scopus</i>	Nº documentos ausentes no <i>WoS</i>	Nº documentos ausentes no <i>Scopus</i>
(A1) Ausência do periódico, série, <i>proceedings</i> da conferência	138	6	122	3
(A2) Documentos indexados no <i>Scopus</i> e apenas no <i>ISI Proceedings</i>	57	0	20	0
(A3) Apenas referência ao nome dos <i>proceedings</i> da conferência	0	87	0	87
(B1) Erros presentes nos endereços ou data de publicação	14	21	50	41
(B2) Documentos não identificados	3	25	5	33
(B3) Volume, número ou suplemento do periódico ausente	0	47	0	43
(B4) Documentos duplicados	0	0	2	0

O *WoS* tem vindo a sofrer várias alterações ao longo dos últimos anos. A integração do *ISI Proceedings* no *WoS* foi uma das alterações entre muitas outras. Considerando isto e os erros presentes nos endereços ou data de publicação (B1) o número de documentos indexados simultaneamente em ambas as bases de dados é agora superior.

Os valores apresentados na Tabela 5 mostram que a percentagem de documentos referenciados em ambas as bases de dados continua a ser muito semelhante para as duas universidades. A percentagem do total de documentos simultaneamente nas duas bases de dados é de 74% do total de documentos encontrados para a Universidade de Coimbra e 76% para a Universidade de Lisboa.

A percentagem de documentos presentes agora em apenas uma das bases de dados sofreu uma diminuição com o intervalo de variação mais amplo e situado entre 10% e 14%.

Tabela 5. Documentos indexados em 2006 para a Universidade de Coimbra e Universidade de Lisboa, após a correção para possíveis erros na indexação (B1) e alterações da estrutura da base de dados (A2).

Documentos indexados	Universidade	
	Universidade de Coimbra	Universidade de Lisboa
Apenas no <i>Scopus</i>	141	127
Apenas no <i>WoS</i>	165	166
Simultaneamente em <i>Scopus</i> e <i>WoS</i>	849	924
Total	1155	1217

A indexação dos documentos publicados pelas duas universidades foi analisada sem considerar a área científica. Tratando-se de universidades multidisciplinares é importante

estudar a indexação destes documentos por áreas científicas. O *Scopus* classifica os seus periódicos em 4 grandes áreas e num conjunto relativamente grande de categorias. É para o conjunto das 4 áreas científicas que é apresentada, na Tabela 6, a distribuição dos documentos indexados no *Scopus* para a Universidade de Coimbra e de Lisboa.

Tabela 6. Distribuição do número de documentos indexados no *Scopus* por área científica para a Universidade de Coimbra e Universidade de Lisboa em 2006.

Área científica	Universidade de Coimbra		Universidade de Lisboa	
	Documentos únicos no <i>Scopus</i>	Documentos comuns ao <i>Scopus</i> e <i>WoS</i>	Documentos únicos no <i>Scopus</i>	Documentos comuns ao <i>Scopus</i> e <i>WoS</i>
Ciências Físicas	137	465	102	448
Ciências da Saúde	50	57	49	88
Ciências Sociais	10	38	28	25
Ciências da Vida	16	197	18	249
Multidisciplinar	0	0	0	3

Este exercício apenas foi efetuado para os documentos indexados no *Scopus*, porque no momento da análise a classificação disponibilizada pelo *WoS* para os seus periódicos era apenas por categorias (mais de 200 categorias).

A classificação por categorias permite obter conjuntos de periódicos mais homogêneos do que os obtidos ao nível de grandes áreas. No entanto a interpretação dos resultados a este nível pode tornar-se mais complexa. A distribuição dos documentos pelas grandes áreas permite obter uma imagem geral e que é a mais adequada quando se está perante instituições de investigação multidisciplinares.

Os resultados da Tabela 6 mostram que as ciências físicas é a área em que a percentagem de documentos indexados simultaneamente nas bases de dados é maior. É também para esta área que se regista uma maior percentagem de documentos não indexados no *WoS*, cerca de 65% e 52% para a Universidade de Coimbra e Universidade de Lisboa, respetivamente (valores calculados relativamente à Tabela 3).

Os resultados obtidos nesta secção sugerem que na realidade as diferenças nas contagens dos documentos indexados em qualquer uma das bases de dados são inferiores ao inicialmente previsto (antes da identificação dos fatores que justificam as diferenças e que não estão relacionados com as políticas de indexação).

2.4 O impacto no *Scopus* e/ou *WoS* para uma instituição

Considerando as diferenças registadas anteriormente relativamente à contagem de documentos indexados no *Scopus* e *WoS* importa também estudar o impacto da produção científica de cada uma das universidades para os diferentes conjuntos de documentos. Isto permitirá identificar se os documentos contabilizados em apenas uma das bases de dados são de facto importantes ao nível do impacto causado na comunidade. Na Tabela 7 são apresentados os valores obtidos para os diferentes conjuntos de documentos. A média de

citações por documento apresentada na Tabela 7 considera o conjunto de documentos da Tabela 3.

Tabela 7. A média de citações por documento (CPD) obtido para cada um dos conjuntos de documentos indexados em 2006.

Universidade	CPD documentos únicos		CPD documentos em simultâneo		CPD do total de documentos	
	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>	<i>Scopus</i>	<i>WoS</i>
Univ. de Coimbra	0,54	0,85	2,93	3,05	2,50	2,53
Univ. de Lisboa	1,47	1,82	3,57	3,75	3,30	3,12

Os valores determinados para o número médio de citações por documento permite concluir que os documentos indexados no *WoS* no geral possuem maior impacto do que os documentos indexados no *Scopus*. Apenas para o total de documentos indexados no *Scopus* da Universidade de Lisboa foi verificado que o impacto médio destes é superior ao obtido pelo total de documentos indexados no *WoS*. Os documentos indexados simultaneamente nas duas bases de dados são aqueles que obtiveram maior impacto. No entanto, apesar de estes serem os mais citados os documentos presentes em apenas uma das bases de dados devem ser estudados pois existe a possibilidade de alguns serem muito citados. Isto levou a um estudo mais detalhado deste conjunto de documentos. A Figura 4 mostra os resultados alcançados.

Dos documentos indexados apenas no *Scopus* para a Universidade de Coimbra e Universidade de Lisboa 75% e 61%, respetivamente, possuem zero citações. As percentagens mantêm-se elevadas quando se considera apenas os documentos indexados no *WoS* (90% para a Universidade de Coimbra e 82% para a Universidade de Lisboa). Neste conjunto está essencialmente o material apresentado em conferências e que no *WoS* são classificados como *meeting abstracts*. A média de citações por documento para os documentos com uma ou mais citações indexados apenas no *Scopus* é de 2,13 e 3,73 para a Universidade de Coimbra e Universidade de Lisboa, respetivamente. Este valor está próximo do valor determinado para o total de documentos indexados no *Scopus*. É importante salientar aqui que o fator políticas de indexação explica a maior percentagem de documentos não indexados no *WoS* e que receberam uma ou mais citações. A média de citações por documento para os documentos apenas indexados no *WoS* é bastante superior à obtida para os documentos indexados apenas no *Scopus* (8,6 e 10,1 para a Universidade de Coimbra e Universidade de Lisboa, respetivamente). Neste conjunto estão essencialmente os documentos que não foram identificados no *Scopus* devido aos erros de transcrição de informação para a base de dados.

Os resultados obtidos não justificam a utilização das duas bases de dados para descrever o impacto do desempenho científico pois para os documentos apenas indexados em uma das bases de dados a média de citações é efetivamente muito baixa quando comparada com a média dos documentos indexados simultaneamente nas duas bases de dados. No momento da escolha de um das bases de dados para efetuar análises bibliométricas os resultados obtidos sugerem que o *WoS* deve ser escolhido. Na realidade esta base de dados possui um menor número de periódicos indexados, mas os resultados obtidos para o impacto mostram que os documentos publicados nos periódicos não indexados no *WoS*, para as universidades usadas como exemplo, possuem um baixo impacto o que não justifica a utilização de ambas as bases de dados. Os resultados do impacto foram decisivos no processo de seleção pois partiu-se do

princípio que no momento da avaliação do desempenho o impacto será a dimensão mais valorizada.

Perante estes resultados optou-se apenas pela utilização da base de dados *WoS* nos estudos apresentados nos capítulos seguintes.

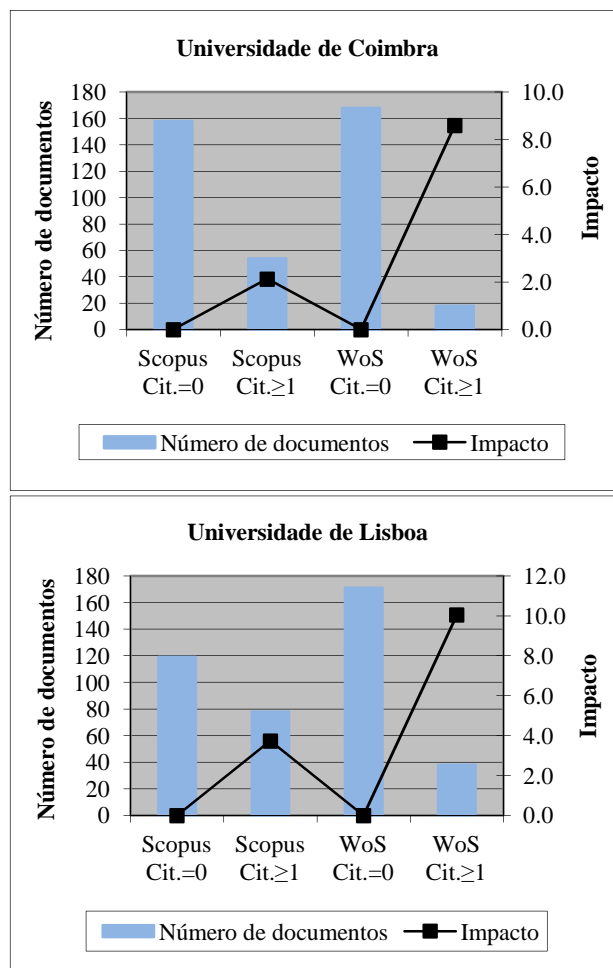


Figura 4. Número de documentos indexados no ano 2006 em apenas uma das bases de dados e respetiva média de citações por documento (impacto).

2.5 Bases de citações que compõem o *WoS*

Para o conjunto dos 210 investigadores holandeses da amostra considerada, as diferenças entre o número de documentos indexados quando se considera apenas as três bases de citações (*SCI-EXPANDED*, *SSCI* e *A&HCI*) e o número de documentos indexados quando se utilizam a globalidade das bases são apresentadas na Tabela 8.

Os resultados, da Tabela 8, mostram que para todas as categorias consideradas existe um incremento do número de documentos contabilizados quando as cinco bases são consideradas na pesquisa. Este aumento é de facto considerável para a categoria *Engineering, Electrical & Electronic*, cerca de 69%, a categoria *Computer Science Theory and Methods* e *Physics Applied* seguem-se com um aumento de 28% e 16% respetivamente. Em áreas mais tecnológicas como as engenharias e as ciências dos computadores a publicação dos resultados das atividades científicas é mais comum sobre a forma de *proceedings* de conferências do que

em periódicos na forma de artigos (Meho e Rogers 2008). Para estas áreas existem conferências de elevado prestígio onde a taxa de aceitação de estudos é muito baixa (Meyer, Choppy, Staunstrup *et al.* 2009) daí a importância dada a este tipo de documentos nestas categorias. A categoria *Oncology* é aquela com o menor incremento, apenas de 0,26%.

A *Engineering, Electrical & Electronic* e *Computer Science Theory & Methods* são aquelas que possuem um dos valores mais baixos para a média de documentos por investigador (28,7 e 29,5 respetivamente), de entre o conjunto de categorias aqui considerado, quando apenas as três bases de citações são consideradas. Na contabilização dos documentos nas cinco bases de citações estas são as categorias que obtêm um maior incremento na média de documentos por investigador.

A percentagem de investigadores que vê o seu número de documentos aumentar quando se consideram as cinco bases de citações é considerável para quase todas as categorias. Exceção é a categoria *Biochemistry Molecular Biology* onde apenas cerca de 7% dos investigadores tem um aumento do número de documentos.

Tabela 8. Número de documentos indexados nas várias bases de citação do WoS para o conjunto de 210 investigadores holandeses no período 2003-2007.

Categoria	Nº investigadores considerados	Total de documentos		Nº investigadores com mais documentos
		SCI-EXPANDED; SSCI; A&HCI	SCI-EXPANDED; SSCI; A&HCI; CPCI-S; CPCI-SSH	
<i>Biochemistry Molecular Biology</i>	44	1520	1530	3
<i>Chemistry, Organic</i>	6	448	460	4
<i>Chemistry, Physical</i>	13	774	809	12
<i>Computer Science, Theory & Methods</i>	20	590	755	19
<i>Engineering, Chemical</i>	5	376	385	4
<i>Engineering, Electrical and Electronic</i>	19	546	922	13
<i>Environmental, Sciences</i>	23	697	722	6
<i>Mathematics, Applied</i>	9	155	162	3
<i>Oncology</i>	10	1525	1529	4
<i>Pharmacology & Pharmacy</i>	14	1441	1448	4
<i>Physics, Applied</i>	18	885	1027	18
<i>Physics, Condensed Matter</i>	16	566	595	12
<i>Radiology, Nuclear Medicine & Medical Imaging</i>	13	1357	1372	5

As implicações da contabilização dos documentos apenas indexados nas bases *SCI-EXPANDED*, *SSCI* e *A&HCI* são consideráveis para algumas categorias se apenas a dimensão quantidade está a ser avaliada sendo necessário um cuidado especial na avaliação dos investigadores destas categorias. No entanto, importa também analisar os efeitos a nível da dimensão impacto. Os resultados desta análise estão mencionados na Tabela 9.

Tabela 9. Número de citações (TC) e número médio de citações por documento (CPD) contabilizadas nas várias bases de citação do WoS para as categorias consideradas, no período 2003-2007 (amostra de 210 investigadores holandeses).

Categoria	TC		CPD	
	SCI-EXPANDED; SSCI; A&HCI	SCI-EXPANDED; SSCI; A&HCI; CPCI-S; CPCI-SSH	SCI-EXPANDED; SSCI; A&HCI	SCI-EXPANDED; SSCI; A&HCI; CPCI-S; CPCI-SSH
<i>Biochemistry Molecular Biology</i>	17389	17434	11,4	11,4
<i>Chemistry Organic</i>	7145	7146	15,9	15,5
<i>Chemistry Physical</i>	6044	6047	7,8	7,5
<i>Computer Science Theory & Methods</i>	2476	2523	4,2	3,3
<i>Engineering, Chemical</i>	2189	2193	5,8	5,7
<i>Engineering, Electrical and Electronic</i>	2515	2626	4,6	2,8
<i>Environmental Sciences</i>	4184	4201	6,0	5,8
<i>Mathematics Applied</i>	413	415	2,7	2,6
<i>Oncology</i>	15828	15828	10,4	10,4
<i>Pharmacology & Pharmacy</i>	5840	5840	4,1	4,0
<i>Physics, Applied</i>	5297	5357	6,0	5,2
<i>Physics, Condensed Matter</i>	4259	4290	7,5	7,2
<i>Radiology, Nuclear Medicine & Medical Imaging</i>	7145	7146	5,3	5,2

O número total de citações não sofre alterações consideráveis quando todas as bases de dados são utilizadas, ao contrário do que foi observado para o número total de documentos. São várias as categorias que apesar de o número de documentos aumentar o número total de citações permanece o mesmo. As categorias em que o número total de citações mais aumenta são a *Engineering, Electrical and Electronic* e *Computer Science Theory and Methods* e *Physics Applied*. Para estas o aumento é de 4,4%, 1,9% e 1,1% respetivamente.

Analisar o aumento do número total de citações quando estamos a utilizar um maior número de bases de citações não é suficiente para inferir sobre as consequências do uso destas. Conclusões mais seguras apenas podem ser obtidas se indicadores independentes da dimensão quantidade forem determinados. O número médio de citações por documento pode ser utilizado para este fim. Os resultados para cada categoria são também apresentados na Tabela 9.

A média de citações por documento permanece a mesma para as categorias *Biochemistry Molecular Biology* e *Oncology* e diminui para as restantes quando são utilizadas todas as bases de citações que compõem o WoS. Esta diminuição é mais acentuada para aquelas categorias que sofreram um maior aumento do número total de documentos e citações. A média de citações por documento diminui cerca de 20%, 38% e 13% para *Computer Science Theory & Methods*, *Engineering, Electrical & Electronic* e *Physics, Applied*. No que respeita ao impacto alcançado os documentos indexados nas bases *CPCI-S*; *CPCI-SSH* parecem ser pouco relevantes para a comunidade científica.

2.6 Conclusões

O estudo da distribuição dos documentos publicados por uma universidade no *Scopus* e *WoS* mostrou que de facto estas bases são complementares. No entanto a análise detalhada dos documentos de duas universidades permitiu concluir que as diferenças nas contagens dos documentos seriam inferiores se os erros de transferência de informação dos periódicos para as bases de dados não estivessem presentes. Numa situação destas as diferenças seriam apenas consequência das diferentes políticas de indexação.

Os resultados alcançados levaram à opção pela base de dados *WoS* nos estudos que serão apresentados posteriormente. O facto de uma elevada percentagem (63%) de documentos indexados apenas no *Scopus* resultarem da ausência dos periódicos no *WoS* poderia levar a que também fosse utilizada a base de dados *Scopus*. No entanto, verificou-se que para estes documentos a média de citações por documento é muito baixa. Valores superiores foram encontrados para os documentos indexados apenas no *WoS*.

A quantidade e o impacto dos documentos publicados por um investigador serão os parâmetros avaliados pelos painéis dos concursos. Considerando que a vertente impacto é a mais importante, terá todo o interesse trabalhar com a base de dados que fornece mais informação sobre o impacto das atividades científicas.

A metodologia de recolha de dados para cada investigador implica que poucos erros, na transferência de dados para as bases de dados, estejam presentes. O *WoS* mostrou ter melhor desempenho neste aspeto. Para cerca de 54% do total de documentos que não foram encontrados no *Scopus* os erros de transferência de informação para a base de dados explicam a ausência destes. Para os documentos apenas indexados no *Scopus* os erros de transferência de informação apenas justificam a ausência, no *WoS*, de 18% do total de documentos.

A análise conduzida relativamente às bases de citações que integram o *WoS* permitiu chegar às seguintes conclusões:

- se o objetivo é medir a quantidade da produção científica, os investigadores das áreas tecnológicas serão os mais prejudicados. Para as restantes áreas apenas uma pequena percentagem dos documentos fica por contabilizar se não são utilizadas as cinco bases de citações;
- pretendendo medir-se o impacto da produção científica a contabilização dos documentos indexados nas cinco bases e das citações resulta num impacto médio que é inferior ao obtido quando apenas são utilizadas as três bases de citações (*SCI-EXPANDED*, *SSCI* e *A&HCI*). Aqui as áreas mais tecnológicas são as que sofrem os maiores decréscimos. Para nenhuma das categorias analisadas se verificou um aumento do impacto médio com a consideração das cinco bases de citações. Este comportamento poderá significar que os documentos apenas indexados no *CPCI-S* e *CPCI-SSH* são menos importantes no que respeita ao impacto;
- os efeitos resultantes da utilização de apenas as bases *SCI-EXPANDED*, *SSCI* e *A&HCI* são opostos em cada uma das dimensões. A utilização de apenas três bases de citações poderá prejudicar os indivíduos das áreas mais tecnológicas na dimensão quantidade, mas irá beneficiar na dimensão impacto. Perante este cenário cabe aos

responsáveis pelo processo de avaliação, no caso de se estar a avaliar investigadores, decidir qual o procedimento que deve ser adotado.

Capítulo 3

Indicadores bibliométricos

Ao longo dos anos vários indicadores têm vindo a ser desenvolvidos com o objetivo de descrever as várias vertentes que caracterizam o desempenho científico. Um conjunto de doze indicadores foi determinado para descrever o desempenho científico dos investigadores que foram utilizados para a definição do modelo apresentado no Capítulo 4.

De entre os indicadores calculados alguns são já utilizados na comunidade científica, tendo sido adaptados para o conjunto de dados disponíveis. Outros indicadores foram alterados no sentido de ultrapassar algumas das suas limitações e novos indicadores foram também desenvolvidos.

O desempenho científico é algo complexo, influenciado por vários fatores e está muito dependente do contexto em que se encontra inserido o objeto avaliado. Atualmente, os indicadores bibliométricos determinados para grandes conjuntos de dados, como por exemplo para a produção científica de um país ou de uma universidade, são aceites como instrumento de medida do desempenho pela comunidade científica. No entanto, a sua aplicação a indivíduos não tem ainda o mesmo nível de aceitação na comunidade científica, apesar da crescente utilização para efeitos de avaliação (exemplos: Costa e Oliveira 2012; Haeffner-Cavaillon e Graillot-Gak 2009).

Nos concursos para Professor Catedrático e Associado aqui considerados é pedido ao júri para avaliar o desempenho científico dos candidatos, mas pouco é dito acerca dos critérios que devem ser utilizados. Assim, é esperado que o júri utilize os seus próprios critérios de acordo com o que este considera ser relevante na avaliação do desempenho científico. Dado que não são conhecidos os critérios aplicados optou-se por selecionar um conjunto de indicadores que descrevem o desempenho científico dos investigadores nas suas várias dimensões. Atendendo ao elevado número de indicadores bibliométricos disponíveis foi necessário proceder à escolha de apenas alguns. Todos os indicadores apresentam as suas vantagens e limitações tornando o processo de seleção ainda mais difícil. Neste estudo procuraram-se indicadores que, para além de descreverem o desempenho científico de investigadores, possam ter estado implícitos nas avaliações efetuadas pelos pares, e permitem a comparação de indivíduos que trabalham em diferentes áreas científicas.

No presente capítulo são abordados os indicadores utilizados na caracterização do desempenho científico.

3.1 Amostra

Amostra de periódicos

Alguns dos periódicos classificados no WoS em pelo menos uma das categorias *Chemistry, Physical (CP)*, *Chemistry, Organic (CO)* e *Chemistry, Multidisciplinary (CM)*, *Materials Science, Multidisciplinary (MSM)*, *Chemistry, Applied (CA)* e *Engineering, Chemical (EC)* e indexados no WoS no período 1990 e 2008 foram utilizados nas várias etapas necessárias para o cálculo do indicador impacto relativo do periódico (*JRI*) que será apresentado em seguida.

Optou-se pela escolha de periódicos classificados em categorias que apresentam culturas de citação distintas entre si.

Amostra de investigadores

Um conjunto de 34 investigadores holandeses foram utilizados nas sucessivas etapas que envolveram o cálculo dos indicadores h_{nf} , NIR , NA_m que serão apresentados ao longo deste capítulo. A produção científica utilizada foi aquela que se encontra indexada no *WoS* durante o período 2000-2007. Os investigadores selecionados respeitam os seguintes critérios:

- possuem como afiliação uma instituição de ensino superior holandesa;
- são os investigadores holandeses com um maior número de documento nas categorias selecionadas (*Pharmacology & Pharmacy* e *Mathematics, Applied*) no período 2000-2007 e têm pelo menos 50% dos documentos publicados numa destas categorias;
- o número de documentos publicados neste período é superior a 10. Este limite foi estabelecido para evitar um número muito baixo de documentos por investigador.

Optou-se pela seleção de investigadores com pelo menos 50% dos documentos publicados em *Pharmacology & Pharmacy* e *Mathematics Applied*, pois trata-se de categorias com diferentes culturas de citação e publicação. Isto permitirá mostrar os efeitos da normalização considerada nos indicadores.

A justificação pela escolha de investigadores holandeses foi já enunciada no Capítulo 2.

3.2 Metodologia

O conjunto de indicadores selecionado relaciona-se com os seguintes tópicos:

- quantificação da produção científica;
- impacto dos periódicos onde os investigadores publicaram;
- impacto da produção científica dos investigadores;
- prestígio da instituição que constitui a afiliação dos autores;
- colaboração.

3.2.1 Cálculo dos indicadores

3.2.1.1 Quantificação da produção científica

A quantificação da produção científica é feita, normalmente, utilizando como indicador o número total de documentos indexados em bases de dados multidisciplinares como o *Scopus* e o *WoS*. Aqui optou-se pela utilização de um indicador que tem em consideração o número de autores de cada documento. O tratamento de documentos com multiautoria é um assunto amplamente discutido na bibliometria. A crescente especialização e a sofisticação das atividades de investigação de hoje exigem que os investigadores procurem colaborações. A colaboração entre grupos de investigação permitirá a fertilização de ideias e uma complementaridade de bases de conhecimentos e capacidades técnicas. Dito de um modo diferente cada grupo irá enriquecer as suas linhas de pesquisa incorporando o conhecimento partilhado. Se cada grupo contribui com o seu conhecimento para a resolução de um

problema (o que não o conseguiria fazer se trabalhasse isolado) parece razoável utilizar um indicador que considera o número de autores por documento.

Num documento com vários autores não é comum mencionar a contribuição de cada autor para o trabalho desenvolvido. Para algumas áreas científicas a posição do autor na lista de autores fornece alguma indicação sobre a relevância deste para o trabalho apresentado. Em outras áreas os autores são simplesmente mencionados por ordem alfabética (Frandsen e Nicolaisen 2010; Waltman 2012). Existem várias propostas para a contagem dos documentos publicados por um investigador em que a lista de autores é tida em consideração:

- Cole e Cole 1973 sugerem que num documento com vários autores este deve ser atribuído apenas ao primeiro autor.
- Outros autores propõem a atribuição de uma fração do documento de acordo com a posição do autor na lista de autores (Hagen 2008; Hodge e Greenberg 1981; van Hooydonk 1997).
- Burrell e Rousseau 1995 sugerem que cada autor deve receber $\frac{1}{a}$ documentos sendo a o número de autores em cada documento.

- ***Número de documentos fracionados (Number of Documents Fractioned, NDF)***

A escolha da metodologia que deve ser utilizada não é simples porque o indicador será determinado para um conjunto de investigadores que trabalham em diferentes áreas científicas. Perante isto optou-se pela seleção de uma metodologia que trata de igual modo as listas de autores das diferentes áreas (Burrell e Rousseau 1995). No entanto, a aplicação de outras metodologias é algo que deverá ser explorado em trabalhos futuros.

Um documento p com a_p autores foi contabilizado com um peso de $\frac{1}{a_p}$ para o total de documentos publicados por um investigador. A produção total de um investigador é então calculada como:

$$NDF = \sum_{p=1}^P \frac{1}{a_p} \quad (1)$$

O parâmetro P é o número total de documentos que pertencem ao investigador e que estão indexados no WoS.

3.2.1.2 Impacto dos periódicos

- ***Source Normalized Impact per Paper (SNIP) e SCImago Journal Rank (SJR)***

Em geral os investigadores procuram publicar os resultados do seu trabalho nos periódicos que consideram de maior prestígio e mais impacto na área respetiva.

O esforço requerido a um investigador para publicar os resultados dos seus estudos em um dado periódico depende, em geral, do prestígio de cada periódico. Isto justifica a utilização do prestígio e impacto dos periódicos onde os investigadores publicaram. Vários indicadores têm sido desenvolvidos para este fim. Neste trabalho foram utilizados o *SNIP*

(Moed 2010) e o *SJR* (Gonzalez-Pereira, Guerrero-Bote e Moya-Anegon 2010). Ambos os indicadores contabilizam as citações obtidas pelos documentos publicados por um dado periódico, mas o *SJR* vai mais longe considerando no seu cálculo o impacto do periódico citante.

O *SNIP* e *SJR* sofreram algumas alterações recentes (Scopus 2013a; Waltman, van Eck, van Leeuwen *et al.* 2013) no modo como são determinados. Dado que estas alterações aconteceram num período posterior à realização do trabalho aqui apresentado os indicadores são descritos utilizando a fórmula de cálculo inicial.

Para um periódico *i* o *SNIP* para 2007 dos documentos publicados em 2004-2006 é calculado utilizando a seguinte expressão:

$$SNIP_i = \frac{RIP_i}{RDCP_i} \quad (2)$$

onde RIP_i é o número médio de citações por documento obtidas em 2007 pelos documentos publicados em 2004-2006 do periódico *i*. O $RDCP_i$ é o potencial relativo da base de dados que é calculado pela expressão:

$$RDCP_i = \frac{DCP_i}{DCP_m} \quad (3)$$

Nesta expressão o *DCP* representa o número médio de referências activas por documento citante de todos os documentos citantes publicados em 2007 e que citam a revista *i*. As referências ativas são os documentos publicados em 2004-2006 que são mencionados nas listas de referências dos documentos citantes. O DCP_m representa o *DCP* do periódico que se encontra na posição mediana na base de dados.

O *SJR* do periódico *i* é calculado com a seguinte expressão:

$$SJR = c \frac{PSJR_i}{Art_i} \quad (4)$$

$$PSJR_i = \underbrace{\frac{(1-d-c)}{N}}_1 + \underbrace{e \frac{Art_i}{\sum_{j=1}^N Art_j}}_2 + \underbrace{d \left[\sum_{j=1}^N C_{ji} \frac{PSJR_j}{C_j} CF + \frac{Art_i}{\sum_{j=1}^N Art_j} \sum_{k \in DN} PSJR_k \right]}_3 \quad (5)$$

A parcela 1 representa o prestígio que advém do facto de o periódico estar indexado na base de dados. A variável *N* representa o número total de periódicos indexados na base de dados e as variáveis *d* e *e* são constantes (0,9 e 0,0999 respetivamente).

A parcela 2 representa o prestígio atribuído pela quantidade de documentos indexados. A variável Art_i representa o número de artigos, artigos de revisão e *proceedings paper*

publicados no periódico i , e Art_j o total de artigos, artigos de revisão e *proceedings paper* indexados na base de dados.

A parcela 3 representa o prestígio proveniente das citações dos periódicos que citam i . A variável C_{ji} é o número de citações do periódico j em um dado ano aos documentos do periódico i publicados num período de três anos. A variável C_j é o número de referências do periódico j . O número de referências do periódico j ao conjunto de periódicos indexados na base de dados *Scopus* (universo) será inferior ao total de referências que constam em j para a maioria dos casos. Segundo os autores isto implica a não convergência do sistema. O fator CF foi introduzido para resolver este problema. O objetivo é distribuir entre os periódicos do universo o prestígio que advém da citação de j a outros periódicos fora do universo. Esta distribuição é proporcional ao prestígio acumulado por cada periódico do universo. O CF é calculado utilizando a seguinte expressão:

$$CF = \frac{1 - (\sum_{k \in DN} PSJR_k)}{\sum_{h=1}^N \sum_{k=1}^N C_{kh} \left(\frac{PSJR_k}{C_k} \right)} \quad (6)$$

O numerador representa a quantidade de prestígio disponível (representa o que os autores chama de *dangling nodes*, isto é periódicos que citam periódicos fora do universo) e o denominador a quantidade de prestígio distribuído através das citações alcançadas utilizando uma janela de citações de três anos.

O prestígio acumulado pelos periódicos que não citam outros periódicos é também considerado:

$$\frac{Art_i}{\sum_{j=1}^N Art_j} \sum_{k \in DN} PSJR_k \quad (7)$$

O $SNIP$ e o SJR foram retirados do *Scopus* para o ano de publicação de cada documento. As distribuições obtidas para o $SNIP$ e SJR para o conjunto de documentos publicados por um investigador são muito assimétricas. A utilização do valor médio para o $SNIP$ e o SJR não constitui assim uma boa medida para representar o impacto e o prestígio dos periódicos onde o investigador publicou. Perante isto, foram determinados os valores medianos.

$$SNIP_m = Md(SNIP) \quad (8)$$

$$SJR_m = Md(SJR) \quad (9)$$

- **Impacto relativo do periódico (Journal Relative Impact, JRI)**

O JRI representa a razão entre o número médio de citações por documento obtido para um dado periódico e o número médio de citações esperadas. O número médio de citações esperado por documento é calculado utilizando todos os documentos publicados em periódicos que pertencem à(s) mesma(s) categoria(s) do periódico para a qual se pretende determinar o JRI .

O *JRI* é um indicador normalizado que considera as diferentes culturas de citação das categorias em que os periódicos estão classificados. No seu cálculo são considerados os documentos publicados num período de cinco anos e na contagem das citações é aplicada uma janela de citação variável.

Considere-se um periódico, a , que no período y publicou n_a documentos e que pertence às categorias $i=1,2,\dots,N$, existindo em cada categoria α documentos (foi utilizada a classificação dos periódicos numa dada categoria fornecida pelo *WoS*). No período x os documentos do periódico a obtiveram C_{yxa} citações. O *JRI* é calculado relativamente à categoria i utilizando a seguinte expressão:

$$JRI_{ayi} = \frac{\frac{1}{n_a} \times C_{yxa}}{\frac{1}{\sum_{\alpha \in i} n_\alpha} \times \sum_{\alpha \in i} C_{yxa}} \quad (10)$$

O termo $\sum_{\alpha \in i} n_\alpha$ representa o total de documentos publicados, em todos os periódicos que pertencem à categoria i durante o período y e $\sum_{\alpha \in i} C_{yxa}$ as citações que estes documentos receberam no período x . A variável y é 5 anos para todos os periódicos e a variável x é igual ou superior a cinco anos dependendo da(s) categoria(s) a que o periódico pertence. Se y e x representam cinco anos então estas variáveis referem-se ao mesmo período. Se x representa seis anos então as citações do periódico a são contabilizadas durante o mesmo período a que y se refere mais um ano extra.

A maioria dos periódicos estão classificados em mais do que uma categoria no *WoS*. Considere-se um periódico que pertence às categorias S1 e S2 e que obtém um elevado *JRI*_{ay1} em S1 mas não em S2. Como avaliar globalmente um periódico relativamente aos valores de *JRI*_{ayi} determinados?

Para um periódico com classificação em duas categorias podemos imaginar três situações:

- o conteúdo do documento publicado está relacionado essencialmente com S1 e será de maior interesse para a comunidade científica que representa S1;
- o conteúdo do documento publicado está relacionado essencialmente com S2 e será de maior interesse para a comunidade científica que representa S2;
- o conteúdo do documento publicado está relacionado com S1 e S2 (documentos multidisciplinares) e será do interesse de ambas as comunidades científicas.

Assumindo-se que do total de documentos, n_a , num periódico a , $\frac{n_a}{N}$ documentos irão ser considerados por cada uma das N comunidades científicas, o número de citações esperadas ($C_{yxa}^E|_{\frac{n_a}{N}}$) para $\frac{n_a}{N}$ será:

$$C_{yxa}^E|_{\frac{n_a}{N}} = \frac{n_a}{N} \left(\frac{1}{\sum_{\alpha \in i} n_\alpha} \sum_{\alpha \in i} C_{yxa} \right) \quad (11)$$

$$C_{yxa|n_a}^E = \frac{n_a}{N} \frac{1}{JRI_{ayi}} C_{yxa} \quad (12)$$

$$C_{yxa|n_a}^E = \frac{C_{yxa}}{N} \frac{1}{JRI_{ayi}} \quad (13)$$

O total de citações esperadas será o somatório de todas as citações esperadas considerando todas as N categorias a que o periódico pertence:

$$C_{yxa}^E = \frac{C_{yxa}}{N} \times \sum_{i=1}^N \frac{1}{JRI_{ayi}} \quad (14)$$

A razão entre as citações obtidas e as esperadas representa o *JRI* para o periódico *a*:

$$JRI_{ay} = \frac{C_{yxa}}{\frac{C_{yxa}}{N} \times \sum_{i=1}^N \frac{1}{JRI_{ayi}}} \quad (15)$$

$$\frac{1}{JRI_{ay}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{JRI_{ayi}} \quad (16)$$

A média harmónica aqui adotada assume como hipótese que um periódico classificado em várias categorias publica documentos em que o conteúdo de cada documento se relaciona com apenas uma categoria e que estão distribuídos uniformemente entre elas.

Na realidade podem existir muitos documentos cujo conteúdo é multidisciplinar e portanto as citações recebidas provêm de uma comunidade científica mais abrangente. Nesta situação o número esperado de citações poderá ser superior ao determinado pela equação 13 e o *JRI* inferior ao determinado pela equação 16.

3.2.1.3 Impacto da produção científica

Normalmente, o impacto que os resultados obtidos nas atividades de investigação têm na comunidade científica é utilizado para ajudar na definição da qualidade do desempenho científico. Aqui o mesmo conceito foi adotado. Indicadores muito diferentes têm sido desenvolvidos para descrever o impacto das atividades científicas. Neste trabalho foram considerados os enumerados em seguida.

- ***Documentos muito citados (Highly Cited Documents, HCD)***

Este indicador traduz a percentagem de documentos publicados por um investigador que estão no Topo 10% dos documentos mais citados em cada categoria.

Para determinar este indicador foram selecionados todos os documentos que mencionam Portugal em pelo menos um dos endereços. Para cada categoria e tipo de documento deste conjunto foi determinado o número mínimo de citações necessários para um documento estar no Topo 10% dos mais citados. As citações foram contabilizadas no período compreendido

entre o ano de publicação e o ano em que foi anunciado o concurso. Com esta informação foi determinado o número de documentos pertencentes a um investigador, P_{10} , no Topo 10% em cada categoria. A fração do total de documentos que se encontra no Topo 10% foi determinada utilizando a seguinte expressão:

$$HCD = \frac{P_{10}}{P'} 100 \quad (17)$$

onde P' representa o número total de documentos publicados em oito anos. Os documentos publicados nos últimos dois anos que precederam o concurso não foram contabilizados dado que o período disponível para a contabilização das citações é muito pequeno.

- ***Documentos citantes (Citing Documents, CD)***

Considerando o total de documentos que citam os documentos de um dado investigador, TCD , foi determinado o número de documentos citantes, CD_{ws} , sem auto-citações. A percentagem de documentos que citam os documentos de um investigador eliminado as auto-citações é dada por:

$$CD = \frac{CD_{ws}}{TCD} 100 \quad (18)$$

Com este indicador pretende-se ter uma ideia do impacto extremo de cada investigador na comunidade científica.

- ***Percentagem de documentos citados (Percentage of Cited Documents, PDC)***

Este indicador fornece a percentagem de documentos de um investigador que obteve citações ao longo do tempo. As citações foram recolhidas para o período de tempo compreendido entre o ano de publicação e o ano em que o concurso para a posição de Professor Associado ou Catedrático foi anunciado.

$$PDC = \frac{P_c}{P} 100 \quad (19)$$

onde P_c é o número de documentos com uma ou mais citações.

- ***O índice h_{nf}***

Este indicador é baseado na teoria que define o índice h , mas foi alterado de modo a eliminar algumas das suas limitações (Vieira e Gomes 2011).

No cálculo do índice h_{nf} existem três passos fundamentais. No primeiro passo cada documento publicado pelo investigador é fracionado pelo respetivo número de autores. No segundo é efetuada a normalização das citações obtidas por cada documento e no terceiro passo o indicador é determinado utilizando o procedimento aplicado no cálculo do índice h .

Contagem fracionada dos documentos

No cálculo do índice h_{nf} cada documento (p) é fracionado pelo número de autores (a_p) que constituem a lista de autores. Os motivos na origem deste procedimento foram mencionados anteriormente no indicador NDF .

Normalização de citações

O processo aqui apresentado envolve a normalização das citações documento a documento.

Considere-se um investigador, que num dado período de tempo, publicou P documentos, cada um com um determinado número de autores $a_1, a_2, \dots, a_j, \dots, a_p$. Desse conjunto de documentos selecionou-se um documento, j , do tipo x , publicado no ano y num periódico que pertence às categorias $i=1, 2, \dots, N$ (foi utilizada a classificação para o tipo de documentos e periódicos numa dada categoria fornecida pelo WoS). Para o documento selecionado o número de citações contabilizadas num dado período, $C_{j(xy i)}$, é normalizado utilizando a seguinte expressão:

$$C_{j(xy i)}^n = C_{j(xy i)} \times \frac{I_{xy}}{\frac{1}{N} \times \left(\sum_{k=1}^N I_{xyk} \right)} \quad (20)$$

O I_{xyk} é a média de citações por documento. Esta média é determinada para os documentos do mesmo tipo, publicados no mesmo ano e no conjunto de periódicos que pertencem à mesma categoria do periódico em que foi publicado o documento selecionado. Com este processo pretende-se que a normalização tenha em consideração os vários tipos de documentos, o ano de publicação e a(s) categoria(s) a que pertencem os periódicos em que foram publicados os documentos. Por sua vez, o I_{xy} é a média de citações por documento para os documentos do mesmo tipo e publicados no mesmo ano. O I_{xy} é calculado do seguinte modo:

$$I_{xy} = \frac{\sum_k M_{xyk} \times I_{xyk}}{M_{xy}} \quad (21)$$

onde M_{xy} é o número de documentos do tipo x publicados no ano y em todos os periódicos independentemente da categoria a que pertencem.

Para efetuar esta normalização foram considerados todos os documentos que mencionavam em pelo menos um dos endereços dos autores um país da UE_15. Existem indicadores que na normalização de citações utilizam o conjunto de todos os documentos indexados na base de dados. King 2004 mostrou no seu estudo que as culturas de publicação e citação diferem de região para região. Isto pode ser explicado pelo contexto institucional. O sistema universitário europeu difere significativamente do sistema dos Estados Unidos da América e portanto os parâmetros que afetam as carreiras científicas dos investigadores não são os mesmos. Perante isto considera-se que a utilização no processo de normalização do

conjunto de documentos publicados pelos países da UE_15 é o mais indicado dado que o objetivo final é a aplicação dos indicadores a investigadores europeus.

Utilizando o processo de normalização descrito o impacto de cada documento (medido pelo número de citações obtidas) é comparado com o impacto obtido pelos documentos do mesmo tipo, publicados no mesmo ano e em periódicos pertencentes à(s) mesma(s) categoria(s). Esta normalização permite a comparação do impacto alcançado pelos documentos, facto que não era possível inicialmente.

Para cada categoria, tipo de documento e ano de publicação, a média de citações por documento foi retirada do *WoS*. As citações foram contabilizadas ao longo do tempo utilizando uma janela de citação que variou entre 10 anos e 1 ano.

A contagem fracionada foi aplicada aos documentos publicados por um investigador, mas não às citações obtidas por cada documento. Alguns estudos mostraram que de facto o número médio de citações aumenta com o número de autores (exemplos: Glanzel e Thijs 2004; Vieira e Gomes 2010). Contudo considera-se que se estamos a utilizar o número de citações obtidas por um documento como instrumento de medida de impacto estas não devem ser fracionadas pelo número de autores pois o impacto de um documento não diminui pelo facto de vários investigadores terem contribuído para o trabalho apresentado.

A auto-citação é também um assunto muito discutido quando se avalia o desempenho científico de investigadores. Nas análises bibliométricas as auto-citações são vistas, por muitos, como um instrumento para inflacionar o número de citações afetando deste modo a validade da avaliação do impacto dos documentos através do uso de citações. Isto levou à realização de estudos com o objetivo de identificar quais as razões que levam os autores a fazer auto-citações. Bonzi e Snyder 1991 encontraram que as principais razões que levam um autor a citar o seu próprio trabalho são as mesmas que justificam as outras citações sendo um processo natural da comunicação científica. As razões encontradas foram as seguintes:

- identificar trabalho relacionado com o conteúdo descrito nos documentos;
- mostrar que o trabalho desenvolvido é baseado em outro já efetuado;
- sugerir que o trabalho citado é relevante para a comunidade científica;
- o trabalho citado pode ser a única fonte de dados disponíveis;

Aksnes 2003 procurou analisar a influência das auto-citações no total de citações alcançadas por um conjunto de documentos em pequenos e longos períodos de tempo. Para um conjunto considerável de documentos (45000) verificou que com uma janela de citação de três anos as auto-citações representam cerca de 30% do total de citações e que este é o período onde o pico máximo é atingido. A partir deste momento diminui a influência das auto-citações. O autor também verificou que o número de auto-citações aumenta com o número de autores por documento. No entanto, este aumento representa uma pequena parcela do total obtido de citações. Conclusão semelhante foi também alcançada por Glanzel e Thijs 2004. Os autores concluíram que um maior número de autores não implica que as auto-citações representem uma maior fração do total das citações relativamente aos documentos com poucos autores.

Identificar o verdadeiro papel das auto-citações é algo complexo tornando difícil definir o modo como estas devem ser tratadas. Isto levou a que as mesmas não fossem excluídas no cálculo do h_{nf} .

Determinação do h_{nf}

O índice h_{nf} é determinado usando o procedimento definido para o índice h . Os documentos de um investigador são ordenados por ordem decrescente de citações, mas agora normalizadas, (C_p^n), como se mostra na Tabela 10.

Tabela 10. Exemplo de cálculo do h_{nf} .

Ordenação dos documentos	a_p	Contagem de documentos	C_p^n	Relação entre citações e número de documentos
1	a_1	$\frac{1}{a_1}$	C_1^n	$C_1^n > \frac{1}{a_1}$
2	a_2	$\frac{1}{a_1} + \frac{1}{a_2}$	C_2^n	$C_2^n > \frac{1}{a_1} + \frac{1}{a_2}$
3	a_3	$\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$	C_3^n	$C_3^n > \frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3}$
...
p	a_p	$\sum_{q=1}^p \frac{1}{a_{(q)}}$	C_p^n	$C_p^n \geq \sum_{q=1}^p \frac{1}{a_{(q)}} = h_{nf}$
p+1	a_{p+1}	$\sum_{q=1}^{p+1} \frac{1}{a_{(q)}}$	C_{p+1}^n	$C_{p+1}^n < \sum_{q=1}^{p+1} \frac{1}{a_{(q)}}$
...
P	a_P	$\sum_{q=1}^P \frac{1}{a_{(q)}}$	C_P^n	$C_P^n < \sum_{q=1}^P \frac{1}{a_{(q)}}$

Aqui $C_1^n > C_2^n > C_3^n \dots > C_P^n$. Um investigador tem um $h_{nf} = \sum_{q=1}^p \frac{1}{a_{(q)}}$ se os primeiros p documentos, $1 \dots p$, têm um número de citações tal que $C_p^n \geq \sum_{q=1}^p \frac{1}{a_{(q)}}$ e os restantes documentos, $P-p$, têm um número de citações tal que $C_p^n \leq h_{nf}$.

O h_{nf} corresponde à contagem máxima do agregado $\sum_{q=1}^p \frac{1}{a_{(q)}}$ de tal modo que este valor é igual ou inferior ao número de citações normalizadas do documento p .

- **Indicador normalizado para investigadores (Normalized Indicator for Researchers, NIR)**

O NIR corresponde ao número médio de citações por documento. As citações obtidas por cada documento são normalizadas utilizando o processo descrito para o indicador h_{nf} . No final é determinada a média para o conjunto de documentos de um investigador utilizando a seguinte expressão:

$$NIR = \frac{1}{P} \sum_{j=1}^P C_{j(xy)}^n \quad (22)$$

3.2.1.4 Prestígio da instituição de afiliação

O prestígio da instituição onde o investigador desenvolve as suas atividades poderá influenciar a avaliação do desempenho científico de um investigador. Como medida do prestígio das instituições utilizou-se o *Normalized Impact (NI)* e o *High Quality Impact (QI)* retirado do *ranking* apresentado anualmente pelo grupo *SCImago* para várias instituições (SCImago 2011). Outros indicadores estão agora disponíveis neste *ranking* e que podem ser testados em estudos futuros.

O *NI* compara o impacto obtido pelos documentos publicados por uma instituição num dado período com o impacto médio mundial. Na normalização são considerados o tipo de documento, o ano de publicação e a categoria a que pertencem os periódicos onde os documentos foram publicados.

O *QI* representa a fração de documentos publicados por uma instituição, num determinado período, em periódicos que pertencem ao primeiro quartil da(s) categoria(s) quando ordenados pelo valor decrescente do *SJR*. Ambos os indicadores são baseados na produção científica indexada no *Scopus*.

3.2.1.5 Colaboração

Nesta última dimensão o objetivo é descrever as práticas de colaboração dos investigadores. Por um lado, considera o número de autores por documento e, por outro, a percentagem de documentos com colaboração internacional. Esta dimensão foi considerada pois a colaboração pode ser vista, num processo de avaliação de indivíduos, como um instrumento utilizado pelos investigadores para melhorar a eficiência das suas atividades de investigação.

- **Número de autores por documento (*Number of Authors per Document, NA_m*)**

O número de autores por documento difere de área científica para área científica (Vieira e Gomes 2010). Existem áreas onde a colaboração com outros investigadores é uma prática comum enquanto que noutras tal não se verifica. Nestas condições a avaliação da colaboração dos investigadores considerando o número de autores por documento deve ser efetuada utilizando indicadores normalizados.

O indicador *NA_m* corresponde ao número de autores por documento (valor mediano do conjunto de documentos de um investigador). O respetivo processo de normalização é apresentado em seguida.

Considere-se um investigador que, num dado período de tempo, publicou *P* documentos, cada um com um determinado número de autores *a₁, a₂, ..., a_j, ..., a_p*. Desse conjunto de

documentos selecionou-se um documento, j , do tipo x , publicado no ano y num periódico que pertence às categorias $i = 1, 2, \dots, N$ (foi utilizada a classificação para o tipo de documentos e periódicos numa dada categoria fornecida pelo *WoS*). Para o documento selecionado o número de autores, $NA_{j(xy)i}$, foi normalizado utilizando a seguinte expressão:

$$NA_{j(xy)i}^n = NA_{j(xy)i} \times \frac{MNA_{xy}}{\frac{1}{N} \times \left(\sum_{k=1}^N MNA_{xyk} \right)} \quad (23)$$

O MNA_{xyk} é a média de autores por documento. Esta média é determinada para os documentos do mesmo tipo, publicados no mesmo ano e no conjunto de periódicos que pertencem à(s) mesma(s) categoria(s) do periódico em que foi publicado o documento selecionado. Com este procedimento pretende-se que a normalização tenha em consideração os tipos de documentos, o ano de publicação e a(s) categoria(s) a que pertencem os periódicos onde foram publicados os documentos. O MNA_{xy} é a média de autores por documento para os documentos do mesmo tipo e publicados no mesmo ano independentemente da categoria. O MNA_{xy} é calculado do seguinte modo:

$$MNA_{xy} = \frac{\sum_k M_{xyk} \times MNA_{xyk}}{M_{xy}} \quad (24)$$

onde M_{xy} é o número de documentos do tipo x publicados no ano y em todos os periódicos independentemente da categoria a que pertencem.

O NA_m é obtido calculando o valor mediano:

$$NA_m = Md(NA_{j(xy)i}^n) \quad (25)$$

No processo de normalização apenas foi utilizado o conjunto de documentos publicados pelos 174 candidatos aos concursos. A normalização poderia ter sido efetuada utilizando o conjunto de documentos publicados pela UE_15, o que implicaria extrair um grande conjunto de documentos do *WoS* para posterior contagem do número de autores. Esta operação seria bastante demorada devido às restrições impostas pelo *WoS* relativamente ao número de registos que é possível extrair de uma só vez (máximo de 500 registos). No entanto, considera-se que o conjunto de documentos utilizados na normalização não afeta a validade das conclusões alcançadas.

- ***Documentos com colaboração internacional (Documents with International Collaboration, DIC)***

Este indicador fornece a proporção do total de documentos publicados pelo investigador com pelo menos uma colaboração internacional.

$$DIC = \frac{P_{IC}}{P} 100 \quad (26)$$

Nesta expressão, P_{IC} denota o número de documentos com pelo menos uma colaboração internacional.

3.2.2 Metodologia de pesquisa no *WoS* para investigadores

A amostra de investigadores holandeses mencionada anteriormente foi utilizada para estudar as propriedades apresentadas pelo h_{nf} , NIR e NA_m . A recolha da produção científica destes investigadores foi efetuada no *WoS* utilizando a metodologia apresentada no Capítulo 2.

Em seguida apresenta-se uma descrição detalhada dos indicadores desenvolvidos (h_{nf} , o NIR , o NA_m e o JRI).

3.3 Indicadores desenvolvidos

3.3.1 Impacto relativo do periódico (*Journal Relative Impact, JRI*)

No desenvolvimento do *JRI* vários aspetos foram analisados pretendendo-se com isto obter um indicador que permite a comparação do desempenho científico de periódicos que pertencem a diferentes domínios científicos.

Os vários aspetos analisados são descritos em seguida.

Tipologia dos documentos

O impacto dos documentos na comunidade científica depende do tipo de documentos e da área científica. Documentos classificados como material editorial, notas, cartas obtêm menos citações do que documentos classificados como artigos, artigos de revisão ou *proceedings paper* para a maioria das áreas científicas. Isto pode ser verificado na Tabela 11 para algumas das categorias utilizadas no *WoS* na classificação dos periódicos.

As médias foram determinadas utilizando os documentos publicados em 2003 em periódicos que estão classificados nas categorias apresentadas. As citações foram contabilizadas num período de cinco anos.

Perante as diferenças observadas no impacto para os vários tipos de documentos, no cálculo do *JRI* apenas foram contabilizados os tipos de documentos responsáveis no mínimo por 5% do total de citações obtidas por todos os documentos de uma categoria num período de 5 anos. Uma percentagem superior a 5% não foi considerada para evitar o uso de um número muito restrito de tipo de documentos. Com este procedimento pretendeu-se uma seleção dos tipos de documentos que está dependente do rendimento dos mesmos em relação às citações e não de alguma ideia preconcebida em relação à relevância das citações e que não é possível generalizar.

Tabela 11. Número médio de citações por documento (*CPD*) para diferentes tipos de documentos e categorias.

Categoria	Tipo de documento	CPD
<i>Biochemistry & Molecular Biology</i>	Artigo	18,5
<i>Biochemistry & Molecular Biology</i>	<i>Proceedings paper</i>	10,3
<i>Clinical Neurology</i>	Artigo	10,9
<i>Clinical Neurology</i>	Material editorial	3,0
<i>Clinical Neurology</i>	Cartas	1,6
<i>Clinical Neurology</i>	Artigos de revisão	19,5
<i>Computer Science, Information Systems</i>	Artigo	5,9
<i>Computer Science, Information Systems</i>	<i>Proceedings paper</i>	5,8
<i>Engineering, Chemical</i>	Material editorial	0,4

Tipo de normalização

Existem na literatura diferentes processos de normalização que tentam capturar os diferentes aspetos das culturas de citação. A normalização pode ser efetuada utilizando o

processo *cited-side* ou o *citing-side*. No *cited-side* para um conjunto de documentos publicados em periódicos pertencentes, por exemplo, à área da química a média de citações é comparada com a média de citações de todos os documentos que foram publicados em periódicos que pertencem à área da química. Este procedimento permite a comparação dos valores relativos obtidos entre documentos pertencentes a diferentes áreas. No *citing-side* a correção para as diferenças entre as áreas científicas é efetuada contabilizando o comprimento das listas de referências que aparecem nos documentos citantes. De modo a tornar mais clara a definição aqui apresentada para o processo *cited-side* e *citing-side* um exemplo é demonstrado para cada caso.

Cited-side

Considere-se um documento, j , do tipo x , publicado no ano y num periódico que pertence à área i . Para o documento selecionado o número de citações contabilizadas, $C_{j(xy i)}$, num dado período são normalizadas utilizando a seguinte expressão:

$$C_{j(xy i)}^n = \frac{C_{j(xy i)}}{I_{xy i}} \quad (27)$$

O $I_{xy i}$ é a media de citações por documento da área i . Esta média é determinada para os documentos do mesmo tipo, publicados no mesmo ano e no conjunto de periódicos que pertencem à área i .

Citing-side

Este processo de normalização foi proposto por Pinski e Narin 1976, mas só recentemente começou a ser explorado na comunidade científica (Zitt e Small 2008) a sua utilização no desenvolvimento de indicadores.

Considere-se um conjunto de documentos P , publicado por um periódico b , no período z , e que é citado pelo periódico i no ano m . O número médio de citações por documento, CPD do periódico b é normalizado usando a seguinte expressão:

$$CPD = \frac{1}{P_{bz}} \sum_i C_{ibmz} \times w_{imz} \quad (28)$$

onde C_{ibmz} é o número de citações do periódico i no ano m aos documentos do periódico b publicados no período z .

O w_{imz} é determinado pela seguinte expressão:

$$w_{imz} = \frac{m_{smz}}{m_{imz}} \quad (29)$$

onde m_{imz} é o número médio de referências do periódico i por documento a documentos publicados entre $[m-1...m-z]$ e m_{smz} é o número médio de referências por documento a todos os documentos publicados entre $[m-1...m-z]$. As referências consideradas nesta normalização são chamadas de referências ativas. As restantes referências não são contabilizadas no processo de normalização.

Neste tipo de normalização as citações recebidas por um periódico são pesadas na proporção inversa do número médio de referências ativas por documento dos periódicos citantes. Existem várias formas de cálculo de w_{imz} (Leydesdorff e Opthof 2010; Moed 2010). Aqui apenas foi apresentada a mencionada por Zitt e Small 2008 para introduzir o conceito de *citing-side*.

No cálculo do *JRI* foi utilizada a normalização *cited-side*. O *WoS* e *Scopus* fazem a classificação dos periódicos em áreas científicas a vários níveis, no entanto várias críticas são apresentadas relativamente às classificações definidas. Estabelecer as fronteiras para cada categoria é algo complicado. Com a classificação dos periódicos pretende-se obter categorias que sejam o mais homogêneas possível. No entanto, a avaliação do grau de homogeneidade de cada categoria é complexa. Isto representa a principal desvantagem deste tipo de normalização. As limitações deste tipo de normalização seriam, possivelmente, ultrapassadas se uma classificação documento a documento estivesse disponível nestas bases de dados. A principal vantagem associada à normalização *citing-side* é a independência relativamente a um esquema de classificação dos periódicos ou documentos em áreas científicas. A normalização *citing-side* também apresenta limitações. Neste tipo de normalização o crescimento da literatura numa dada área científica não é considerado (Moed 2010; Waltman, van Eck, van Leeuwen *et al.* 2013; Zitt e Small 2008). O número de publicações indexadas nas bases de dados tem aumentado ao longo dos anos não só pelo maior número de periódicos indexados, mas também pelo crescimento da produção científica a que se tem assistido nos últimos anos.

Apesar da dependência de um esquema de classificação da normalização *cited-side*, Glanzel, Schubert, Thijs *et al.* 2011 mostraram que este tipo de normalização permite obter resultados muito próximos dos obtidos utilizando indicadores baseados na normalização *citing-side*. van Eck e Waltman 2012 mostraram que a normalização *cited-side* é ligeiramente inferior à normalização *citing-side*, mas que estas diferenças podem ser diminuídas consideravelmente através da utilização de um esquema de classificação mais adequado.

Os resultados dos poucos estudos realizados com o objetivo de mostrar as vantagens e limitações da normalização baseada no *cited-side* e *citing-side* ainda não são suficientes para apontar qual a melhor metodologia.

Definição da janela de citação

Após a definição do tipo de normalização a utilizar torna-se necessário definir a janela de citação. Vários aspetos têm que ser definidos: a metodologia que será utilizada na contagem das citações e o comprimento da janela de citação. Estes aspetos serão analisados em seguida.

A evolução do número médio de citações por documento ao longo do tempo foi estudado para as duas categorias apresentadas na Figura 5.

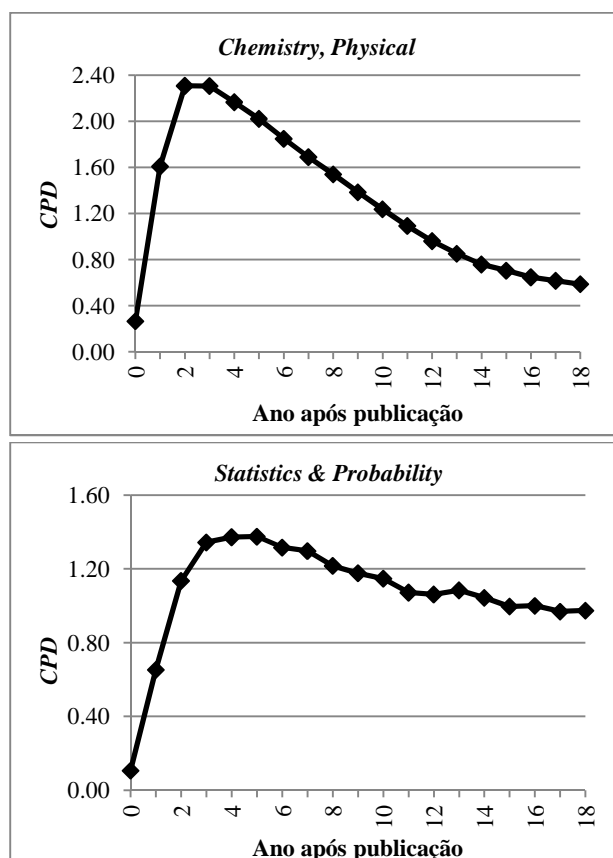


Figura 5. Evolução do número médio de citações por documento (CPD) em *Chemistry, Physical* e *Statistic & Probability* para os documentos publicados entre 1990 e 2008.

Para os periódicos classificados no WoS nas categorias *Chemistry, Physical* e *Statistic & Probability* foi contabilizado o número de documentos publicados entre 1990 e 2008 e o número de citações obtidas em cada ano. O número de citações por documento apresentado para o 4º ano foi calculado considerando o número de citações obtidas por todos os documentos publicados entre 1990-2004 no quarto após o ano de publicação.

A evolução do número de citações por documento sugere que este tipo de estudo deveria ser efetuado para um período mais longo. Após o 18º ano as citações obtidas podem contribuir consideravelmente para o total de citações. Contudo é importante ter em consideração que os hábitos de publicação e citação têm sofrido alterações ao longo dos anos. Atualmente os investigadores publicam e citam mais do que no passado. Por um lado existe mais financiamento disponível para a investigação, o número de investigadores tem aumentado ao longo dos anos (Bettencourt, Kaiser e Kaur 2009), os avanços tecnológicos têm permitido aos investigadores desenvolverem mais trabalho num espaço de tempo menor e as colaborações têm-se intensificado. Disto resulta que a produção científica tem aumentado ao longo das últimas décadas (Board 2012; Larsen e von Ins 2010) disponibilizando à comunidade científica um maior número de referências que servem de suporte ao desenvolvimento das atividades científicas. Por outro lado o número de periódicos indexados no WoS tem vindo a ser incrementado ao longo dos anos (Reuters 2012c). Estes fatores poderão justificar os elevados valores ainda obtidos para o 18º ano.

Na Figura 5 é possível observar que para a categoria *Chemistry, Physical* o pico máximo é obtido entre o 2º e o 3º ano após publicação e o valor médio de citações por documento é de

2,30. Para a categoria *Statistic & Probability* o pico máximo é atingido mais tarde, entre o 4º e o 5º ano e o valor médio de citação por documento é de 1,40. Após atingir o pico máximo verifica-se um decréscimo mais acelerado na categoria *Chemistry, Physical* da média de citações por documento. Para a categoria *Statistic & Probability* o número médio de citação por documento no 18º é agora superior ao obtido para a *Chemistry, Physical*.

Perante o comportamento observado optou-se por definir uma janela de citação que permita a comparação justa entre diferentes áreas científicas. Assim o comprimento da janela de citação utilizada será variável de modo a considerar mais ou menos a mesma percentagem do total de citações que são esperadas para os documentos de cada categoria após 18 anos desde a data de publicação. Aqui considerou-se que a percentagem deveria rondar os 18%. Para chegar a este valor considerou-se várias categorias com diferentes culturas de citação e os documentos publicados durante um período de 5 anos. Para estes documentos determinou-se a razão entre as citações obtidas em 5 anos e as obtidas em 19 anos (incluindo o ano de publicação). O maior valor obtido foi o selecionado garantindo deste modo que em nenhuma das categorias esta percentagem seria atingida em menos de 5 anos. Foram considerados os documentos publicados em 5 anos pois existem periódicos que publicam um baixo número de documentos por ano. Quanto mais baixo é este número mais difícil é interpretar os resultados fornecidos pelos indicadores bibliométricos.

As citações podem ser contabilizadas ao longo de vários anos e neste caso diz-se que o método diacrónico está a ser aplicado. No método síncronico as citações apenas são contabilizadas durante um único ano, isto é, o ano citante é fixo (Glanzel 2004; Ingwersen, Larsen, Rousseau *et al.* 2001). Nas Tabelas 12 e 13 é mostrada a contagem de citações em cada um dos métodos.

Tabela 12. Contabilização das citações utilizando a metodologia diacrónica.

Ano de publicação	Ano de citação					TC
	2007	2008	2009	2010	2011	
2007	5	8	9	12	16	50
2008	-	8	10	12	10	40
2009	-	-	4	3	8	15
2010	-	-	-	2	1	3
2011	-	-	-	-	1	1

Tabela 13. Contabilização das citações utilizando a metodologia síncronica.

Ano de publicação	Ano citação 2011
2007	16
2008	10
2009	8
2010	1
2011	1
TC	36

No desenvolvimento de indicadores de desempenho para periódicos o tamanho da janela de citação é algo que deve ser definido cuidadosamente. Na definição desta janela existem pré-requisitos contraditórios. Por um lado pretende-se um indicador que providencie

informação sobre o desempenho num curto período de tempo e isto apenas é possível utilizando uma janela de citação pequena (entre 2 a 3 anos). Por outro lado, pretende-se que o indicador represente de igual modo as áreas científicas com pequenos e longos períodos de maturação das citações e isto apenas é possível utilizando longas janelas de citação. A utilização de pequenas e longas janelas de publicação também está relacionado com a previsão do impacto. Indicadores com algum poder de previsão são necessários. Perante isto estudou-se a influência da metodologia sincrónica e diacrónica e de janelas de citação com diferente comprimento na previsão do impacto para periódicos classificados na categoria *Chemistry, Physical* e outros na categoria *Statistic & Probability*.

Para a categoria *Statistic & Probability* foi determinada a média de citações para os documentos publicados no período 1990-1992 contabilizando as citações no período 1990-1992 (metodologia diacrónica e janela de citação pequena) e a média de citações dos documentos publicados em 1990-1994 foi determinada contabilizando as citações no período 1990-1996 (metodologia diacrónica, janela de citação longa e que corresponde ao tempo necessário para atingir 18% do total de citações esperadas). Para os documentos publicados em 1990-1992 foi ainda determinada a média de citações obtidas em 1993 (metodologia sincrónica e janela de citação pequena) e em 1996 (metodologia sincrónica e janela de citação longa). Os valores determinados para cada periódico foram comparados com os esperados para um período de observação de citações mais longo (1990-2008). Os resultados são apresentados na Figura 6. Nesta figura é possível avaliar os coeficientes de *Pearson* para dois aspetos: (1) metodologia utilizada na contagem de citações; (2) período de contagem de citações.

Para ambas as metodologias os resultados mostram que o coeficiente de correlação de *Pearson* é superior quando são aplicadas janelas de citação mais longas (superior em 10% na metodologia diacrónica e em 7,6% na metodologia sincrónica). Os coeficientes na metodologia diacrónica são superiores em 2% em relação à metodologia sincrónica para longas janelas de citação. Para pequenas janelas de citação a diferença obtida é mínima. O coeficiente obtido para a metodologia diacrónica é superior ao obtido para a metodologia sincrónica em cerca de 0,14%.

A mesma análise foi efetuada utilizando a categoria *Chemistry, Physical*. Trata-se de uma categoria que é caracterizada por um período de maturação mais rápido do que o da *Statistic & Probability* e portanto é também de todo o interesse estudar a influência do tamanho da janela de citação e da metodologia de contagem de citações na previsão do impacto.

Para a categoria *Chemistry, Physical* foi determinada a média de citações para os documentos publicados no período 1990-1992 contabilizando as citações no período 1990-1992 (metodologia diacrónica e janela de citação pequena) e a média de citações dos documentos publicados em 1990-1994 foi determinada contabilizando as citações no período 1990-1994 (metodologia diacrónica, janela de citação longa e que corresponde ao tempo necessário para atingir 18% do total de citações esperadas). Para os documentos publicados em 1990-1992 foi ainda determinada a média de citações obtidas em 1993 (metodologia sincrónica e janela de citação pequena) e em 1994 (metodologia sincrónica e janela de citação longa). Os valores determinados para cada periódico foram comparados com os esperados para um período de observação de citações mais longo (1990-2008). Os resultados obtidos são apresentados na Figura 7. Na categoria *Chemistry, Physical* são observadas diferenças superiores. Também para esta categoria os coeficientes de correlação de *Pearson* são

superiores quando janelas de citação mais longas são aplicadas, mas as diferenças são superiores às observadas na *Statistic & Probability* (superior em 13% na metodologia diacrónica e em 14% na metodologia síncrona). Para a metodologia síncrona observa-se que os coeficientes são superiores em relação à diacrónica, comportamento que é oposto ao verificado na *Statistic & Probability*. Os coeficientes na metodologia síncrona são superiores em 13% e 15% em relação à metodologia diacrónica. O maior valor (15%) foi encontrado para as longas janelas de citação.

Os coeficientes de *Pearson* obtidos foram no geral superiores para a *Statistic & Probability*. Uma possível explicação para isto poderá ser a interdisciplinaridade dos periódicos considerados. Para ambas as categorias verificou-se que mais de 50% dos periódicos pertencem a mais do que uma categoria. Se o comportamento relativamente às citações diferir muito entre categorias o poder de previsão será afetado consideravelmente. O maior poder de previsão obtido para a *Statistic & Probability* poderá significar que os periódicos desta categoria estão classificados também em outras categorias com uma cultura de citação próxima desta.

Os valores determinados para os coeficientes evidenciam correlações fortes e significativas ($p < 0,001$).

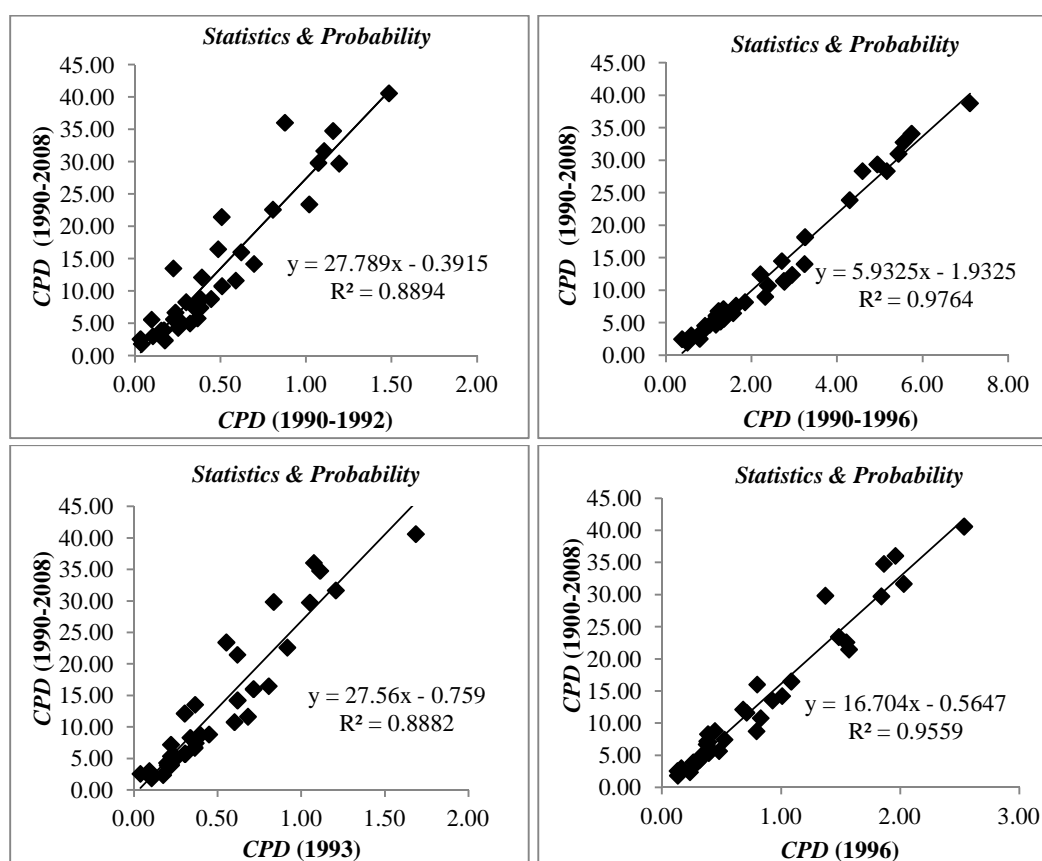


Figura 6. Influência das características da janela de citação na previsão do número médio de citações por documento (CPD) para um conjunto periódicos classificados na categoria *Statistics & Probability*.

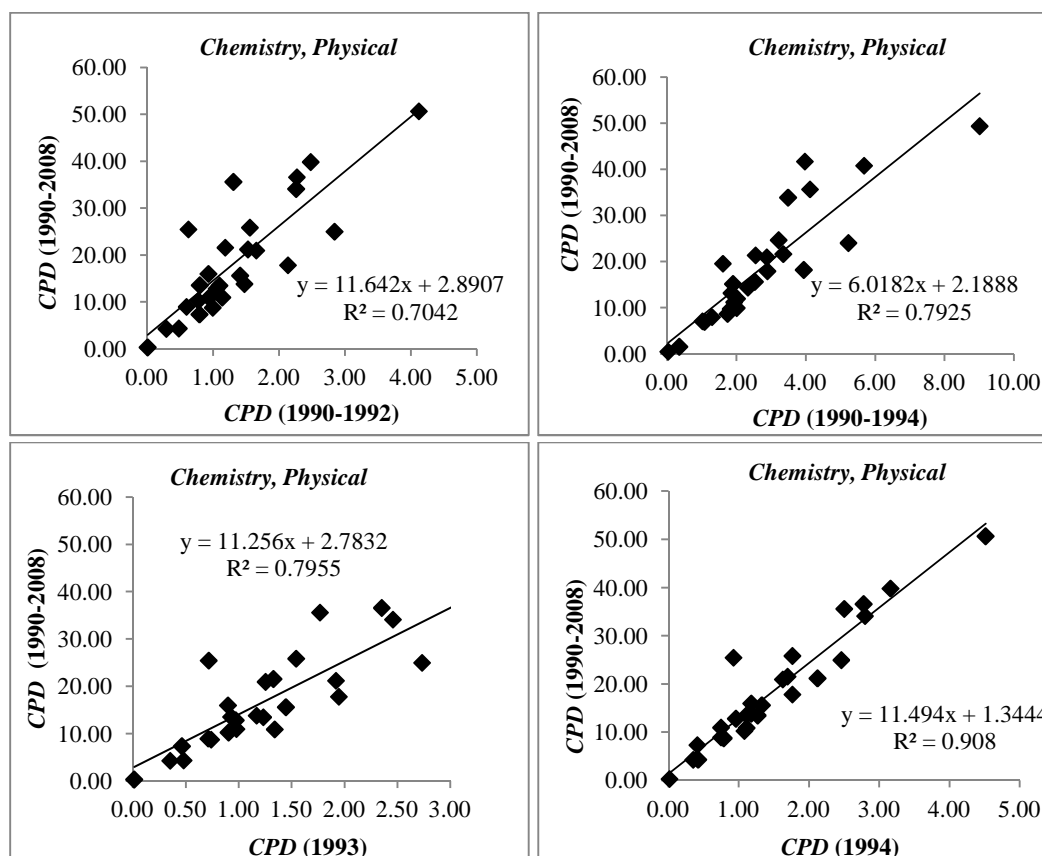


Figura 7. Influência das características da janela de citação na previsão do número médio de citações por documento (CPD) para um conjunto de periódicos classificados na categoria *Chemistry, Physical*.

Os resultados obtidos não permitem uma escolha clara entre a metodologia síncronica e diacrónica, mas são suficientes para concluir que janelas de citação mais longas permitem melhores previsões.

Apesar dos resultados inconclusivos para a seleção da metodologia a utilizar na contagem das citações a metodologia diacrónica foi adotada por se considerar que esta é menos afetada por possíveis flutuações na contagem de citações. Ingwersen, Larsen, Rousseau *et al.* 2001 e Ingwersen, Larsen e Wormell 2000 também sugerem que a contagem de citações deverá ser efetuada utilizando a metodologia diacrónica quando o objetivo é avaliar o desempenho pois este procedimento permite caracterizar melhor o impacto através da contagem das citações ao longo do tempo.

No *JRI* é utilizada uma janela de citação variável de modo a considerar o desfasamento entre o ano de publicação e o ano de citação em cada categoria. Para a categoria *Statistic & Probability* foi identificado que o tempo necessário para atingir cerca de 18% do total de citações esperadas passados 18 anos sobre o ano de publicação para os documentos publicados em 5 anos seria de 7 anos. Para a categoria *Chemistry, Physical* o tempo necessário é inferior, 5 anos. Em seguida é discutido em pormenor a opção por uma janela de citação cujo comprimento varia de acordo com a categoria.

Considere-se um periódico A que pertence a duas categorias, S1 e S2. O conteúdo do periódico é de maior interesse para a comunidade científica de S1 e portanto o impacto obtido

pelos documentos provém essencialmente de S1. Na Tabela 14 é apresentado o comprimento da janela de citação para S1 e S2.

Tabela 14. Comprimento da janela de citação para S1 e S2.

Categoria	Comprimento da janela de citação (18% do total de citações)
S1	6
S2	5

A relação entre o número médio de citações por documento para S1 e S2 aplicando uma janela de citação de cinco e seis anos é a seguinte:

$$\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=6} > \left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=6} \quad (30)$$

$$\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=5} > \left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=5} \quad (31)$$

A relação entre o aumento da média de citações por documento para S1 e S2 quando se transita de uma janela de cinco anos para uma janela de seis anos é:

$$\frac{\left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=6}}{\left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=5}} > \frac{\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=6}}{\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=5}} \quad (32)$$

A relação entre a média de citações por documento esperada para S1 e S2 quando se usa uma janela de citação de seis anos é:

$$\frac{\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=5}}{\left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=5}} > \frac{\left[\frac{1}{\sum_{\alpha \in S2} n_{\alpha}} \times \sum_{\alpha \in S2} C_{yx\alpha} \right]_{x=6}}{\left[\frac{1}{\sum_{\alpha \in S1} n_{\alpha}} \times \sum_{\alpha \in S1} C_{yx\alpha} \right]_{x=6}} \quad (33)$$

Se uma janela de citação de cinco anos é aplicada para as duas categorias e sabendo que o conteúdo publicado no periódico A é de maior interesse da comunidade S1 então é esperado que:

$$[JRI_{yS1}]_{x=5} > [JRI_{yS2}]_{x=5} \quad (34)$$

O $[JRI_{yS1}]_{x=5}$ será também influenciado pelas citações que provém da comunidade científica S2 sobreavaliando o desempenho científico do periódico em relação a S1. O mesmo resultado é esperado quando uma janela de seis anos é aplicada para as duas categorias. O

numerador e o denominador na equação do JRI_{ayi} serão superiores do que na janela de cinco anos, mas o denominador para $[JRI_{yS1}]_{x=6}$ aumentará mais do que o denominador de $[JRI_{yS2}]_{x=6}$ relativamente à janela de citação de cinco anos considerando-se a inequação 32.

A sobreavaliação de $[JRI_{yS1}]_{x=6}$ pelas citações de S2 quando uma janela de citação de seis anos é utilizada, será atenuada por um aumento considerável do número médio de citações por documento em S1. Nesta situação é esperado que:

$$[JRI_{yS1}]_{x=5} > [JRI_{yS1}]_{x=6} \quad (35)$$

A inequação 33 permite concluir que:

$$[JRI_{yS2}]_{x=5} < [JRI_{yS2}]_{x=6} \quad (36)$$

A comparação do JRI_{ayi} usando as diferentes janelas de citação sugere que:

$$[JRI_{yS2}]_{x=6} < [JRI_{yS2}]_{x=5} < [JRI_{yS1}]_{x=6} < [JRI_{yS1}]_{x=5} \quad (37)$$

Com o descrito pretendeu-se mostrar que o JRI de um periódico que pertence essencialmente a uma categoria caracterizada por um baixo número de citações, neste caso S1, pode ser sobreavaliado pelas citações que provêm de uma presença marginal de documentos que pertencem a uma categoria caracterizada por uma elevada densidade de citações, neste caso S2, quando são utilizadas pequenas janelas de citação. Por outro lado o JRI pode ser subavaliado pela presença destes documentos se são aplicadas janelas longas e fixas. Estes efeitos de sobreavaliação e subavaliação indesejados podem ser atenuados utilizando janela de citação longa e variável.

Estabilidade do JRI

O prestígio e a qualidade dos periódicos são importantes no momento de seleção de um para submissão de um trabalho. Para além da escolha de um periódico dedicado à comunicação dos resultados dos trabalhos desenvolvidos na área a que o conteúdo do documento se refere é igualmente importante a seleção de um periódico que apresente um padrão de qualidade o mais próximo possível do conteúdo submetido para publicação. A avaliação da qualidade de um periódico é feita, normalmente, utilizando indicadores baseados no impacto que os documentos publicados tiveram ao longo do tempo. Perante isto uma análise do impacto dos periódicos ao longo dos anos é fundamental para permitir uma seleção mais rigorosa de entre um conjunto de periódicos dedicados à publicação de conteúdos de uma determinada área. Para um investigador interessa também selecionar um periódico que não apresente grandes oscilações ao longo do tempo. Periódicos com um impacto crescente ou estabilizado ao longo dos anos serão os mais interessantes.

Perante o mencionado anteriormente analisou-se a estabilidade do JRI . A estabilidade do $SNIP$ e SJR foi também estudada por se tratar de indicadores que utilizam um processo de cálculo diferente e por serem já utilizados na comunidade científica. Informação relativa ao

SNIP e *SJR*, para um conjunto de 36 periódicos foram retirados do *Scopus* para o ano 2009, 2008, 2007, 2006 e 2005. O *JRI* foi determinado para os documentos publicados em 2000-2004, 2001-2005, 2002-2006, 2003-2007 e 2004-2008. Na Tabelas 15 e 16 apresentam-se os parâmetros estatísticos para a amostra de periódicos.

Tabela 15. Média, desvio padrão e coeficiente de variação para o *JRI*, *SNIP* e *SJR* determinados para cinco períodos.

Periódico	<i>JRI</i>			<i>SJR</i>			<i>SNIP</i>		
	Média	Desvio padrão	Coeficiente de variação	Média	Desvio padrão	Coeficiente de variação	Média	Desvio padrão	Coeficiente de variação
<i>Chemistry of Materials</i>	2,286	0,098	4,3	0,410	0,039	9,6	2,644	0,046	1,7
<i>Journal of Materials Chemistry</i>	1,728	0,109	6,3	0,348	0,057	16,3	2,018	0,159	7,9
<i>Langmuir</i>	1,490	0,026	1,7	0,339	0,031	9,1	1,986	0,051	2,6
<i>Carbon</i>	1,487	0,113	7,6	0,256	0,061	23,8	2,154	0,285	13,2
<i>Carbohydrate Polymers</i>	0,963	0,021	2,1	0,118	0,022	18,6	1,606	0,270	16,8
<i>Topics in Catalysis</i>	1,166	0,127	10,9	0,174	0,009	5,4	1,120	0,050	4,5
<i>Catalysis Today</i>	1,487	0,136	9,1	0,178	0,024	13,4	1,834	0,161	8,8
<i>Reactive & Functional Polymers</i>	1,061	0,107	10,1	0,112	0,022	19,4	0,972	0,099	10,2
<i>Journal of Catalysis</i>	3,268	0,065	2,0	0,312	0,032	10,3	2,946	0,030	1,0
<i>Journal of Supercritical Fluids</i>	1,135	0,095	8,4	0,124	0,012	9,7	1,502	0,078	5,2
<i>Computers & Chemical Engineering</i>	1,155	0,075	6,5	0,084	0,007	7,9	2,308	0,256	11,1
<i>Journal of Membrane Science</i>	1,737	0,132	7,6	0,183	0,024	13,2	2,106	0,372	17,7
<i>Advanced Synthesis & Catalysis</i>	1,929	0,236	12,2	0,348	0,036	10,4	1,532	0,147	9,6
<i>Aldrichimica Acta</i>	3,805	0,539	14,2	0,973	0,449	46,1	3,122	0,555	17,8
<i>Organic Letters</i>	1,821	0,021	1,1	0,372	0,045	12,1	1,906	0,143	7,5
<i>Current Organic Chemistry</i>	1,209	0,157	13,0	0,251	0,023	9,2	1,144	0,093	8,1
<i>Adsorption Science & Technology</i>	0,320	0,028	8,9	0,052	0,005	9,6	0,348	0,068	19,6
<i>Catalysis Letters</i>	0,797	0,108	13,6	0,136	0,008	5,9	1,022	0,051	5,0
<i>Journal of Physical Organic Chemistry</i>	0,609	0,039	6,4	0,108	0,008	7,6	0,624	0,040	6,5
<i>Colloid and Polymer Science</i>	0,616	0,028	4,5	0,105	0,017	16,5	0,786	0,160	20,4
<i>Journal of Porous Materials</i>	0,400	0,136	34,0	0,074	0,004	5,5	0,486	0,074	15,3
<i>Macromolecular Materials and Engineering</i>	0,653	0,017	2,7	0,103	0,013	12,9	0,876	0,132	15,1
<i>Journal of Elastomers and Plastics</i>	0,189	0,043	22,7	0,046	0,002	5,1	0,610	0,076	12,5
<i>International Journal of Adhesion and Adhesives</i>	0,560	0,069	12,3	0,083	0,015	18,4	1,522	0,376	24,7
<i>Journal of Cellular Plastics</i>	0,314	0,042	13,3	0,052	0,006	11,2	0,814	0,139	17,1
<i>Annual Review of Physical Chemistry</i>	1,612	0,168	10,4	1,581	0,256	16,2	5,032	0,752	14,9

Tabela 16. Média, desvio padrão e coeficiente de variação para o *JRI*, *SNIP* e *SJR* determinados para cinco períodos (continuação)

Periódico	<i>JRI</i>			<i>SJR</i>			<i>SNIP</i>		
	Média	Desvio padrão	Coeficiente de variação	Média	Desvio padrão	Coeficiente de variação	Média	Desvio padrão	Coeficiente de variação
<i>Journal of Photochemistry and Photobiology A-Chemistry</i>	0,871	0,048	5,5	0,150	0,009	6,2	1,400	0,087	6,2
<i>Adsorption-Journal of the International Adsorption Society</i>	0,503	0,156	31,0	0,085	0,023	26,6	0,744	0,348	46,8
<i>International Polymer Processing</i>	0,356	0,056	15,7	0,048	0,003	6,0	0,596	0,170	28,5
<i>International Reviews in Physical Chemistry</i>	0,746	0,049	6,5	0,584	0,202	34,6	2,850	0,424	14,9
<i>Computational Materials Science</i>	0,442	0,073	16,5	0,108	0,021	19,2	1,128	0,200	17,7
<i>Computing in Science & Engineering</i>	0,519	0,075	14,5	0,065	0,02	30,3	0,706	0,073	10,3
<i>Inorganic Materials</i>	0,117	0,019	15,9	0,042	0,001	3,5	0,216	0,068	31,5
<i>Polymer Engineering and Science</i>	0,837	0,073	8,7	0,080	0,002	2,9	1,138	0,148	13
<i>Advances in Polymer Technology</i>	0,515	0,032	6,2	0,060	0,006	10,0	0,836	0,073	8,7

Para o conjunto de periódicos apresentado nas Tabelas 15 e 16 os desvios relativamente ao valor médio do *JRI* determinado para cinco períodos atingem, em média, 10,3% do valor desta. Para o *SNIP* os desvios atingem 13,6% e para o *SJR* 13,4%.

O coeficiente de variação foi utilizado para comparar a estabilidade do *JRI* com a estabilidade do *SNIP* e do *SJR*. O teste do sinal (teste não paramétrico) foi aplicado para verificar as seguintes hipóteses:

H_0 : As medianas dos coeficientes de variação são iguais.

H_1 : As medianas dos coeficientes de variação não são iguais (teste bilateral).

Este teste foi realizado duas vezes e os resultados mostraram que de facto as medianas são diferentes ($p=0,03<0,05$). O resultado foi o mesmo para os dois testes. Dado que também se pretendia verificar se a estabilidade do *JRI* era superior à do *SNIP* e *SJR* as seguintes hipóteses foram consideradas também:

H_0 : As medianas dos coeficientes de variação são iguais.

H_1 : A mediana do coeficiente de variação do *JRI* é inferior à mediana do coeficiente de variação do *SNIP* (teste unilateral).

As mesmas hipóteses foram consideradas utilizando o *SJR*.

Os resultados obtidos mostraram que a mediana do coeficiente de variação do *JRI* é significativamente inferior à mediana do coeficiente de variação obtido para o *SNIP* e *SJR* ($p=0,01<0,05$). Mais uma vez o resultado obtido foi o mesmo para ambos os testes. Isto permite concluir que de facto o *JRI* é o mais estável dos três indicadores.

Poder de previsão do JRI

O estudo da estabilidade de um indicador fornece alguma informação sobre o poder de previsão do mesmo. Indicadores relativamente estáveis terão à partida um poder de previsão razoável. Por outro lado, a instabilidade não é obrigatoriamente sinónimo de baixo poder de previsão. Considere-se, por exemplo, um indicador que tem um comportamento monótono crescente ao longo do tempo. Este será um indicador, à partida, instável, mas isto não implica que o seu poder de previsão seja baixo.

Na Figura 8 é mostrada a relação entre os valores de *JRI*, *SNIP* e *SJR*, de diversos periódicos, calculados para dois períodos. O *JRI* foi calculado para os documentos publicados em 2000-2004 e 2004-2008. O *SNIP* e o *SJR* foram retirados do *Scopus* para o ano 2005 e 2009 para o mesmo conjunto de periódicos.

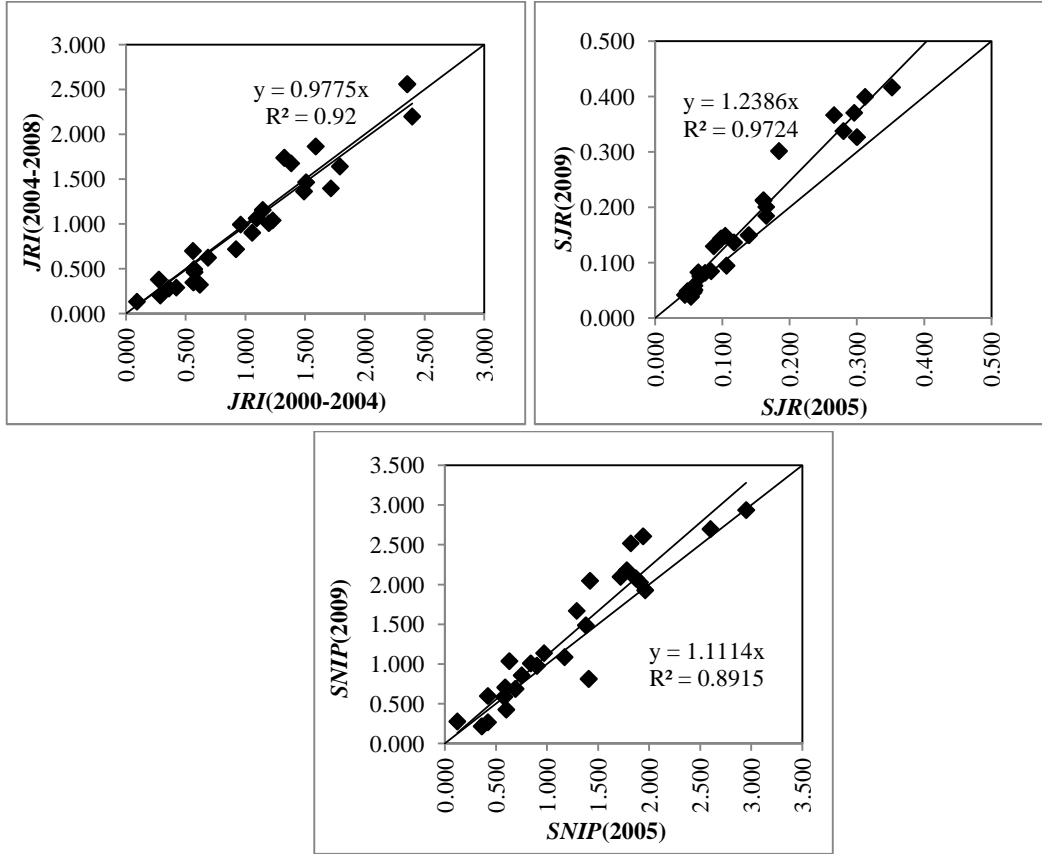


Figura 8. Valores de *JRI*, *SJR* e *SNIP* para um conjunto de periódicos.

Os resultados mostram uma forte correlação entre os valores obtidos nos dois períodos para os três indicadores. As correlações obtidas são significativas ($p < 0,001$).

O *SJR* apesar de ser um indicador mais instável do que o *JRI* apresenta um poder de previsão superior se este for avaliado pelo coeficiente de correlação de *Pearson*. O *SNIP* e *JRI* possuem um poder de previsão similar. O aumento do *SJR* entre os dois períodos é de 24% e para o *SNIP* é de 11%. O mesmo comportamento não se verifica para o *JRI*. O valor do declive encontra-se muito próximo de 1 reforçando a maior estabilidade do indicador relativamente aos restantes.

Para o *SJR* o aumento sugerido pelos resultados poderá justificar-se pelo aumento ao longo dos anos do número de periódicos indexados na base de dados. De modo a ilustrar o efeito deste aspeto no cálculo do *SJR* considere-se a sua fórmula de cálculo.

$$SJR = c \frac{PSJR_i}{Art_i} \quad (38)$$

$$PSJR_i = \underbrace{\frac{(1-d-c)}{N}}_1 + e \underbrace{\frac{Art_i}{\sum_{j=1}^N Art_j}}_2 + d \underbrace{\left[\sum_{j=1}^N C_{ji} \frac{PSJR_j}{C_j} CF + \frac{Art_i}{\sum_{j=1}^N Art_j} \sum_{k \in DN} PSJR_k \right]}_3 \quad (39)$$

Na parcela identificada com o número 3 é efetuado o somatório de todas as citações obtidas pelos documentos publicados no periódico i levando em consideração o tamanho das listas de referências de cada periódico citante e o respetivo prestígio através de $\sum_{j=1}^N C_{ji} \frac{PSJR}{C_j} CF$. Com o aumento do número de periódicos indexados ao longo dos anos a possibilidade dos documentos receberem citações será maior e portanto este somatório irá aumentar ao longo dos anos. Para além disto é ainda adicionado o prestígio acumulado pelos periódicos que não citam outros periódicos proporcionalmente ao número de artigos, artigos de revisão e *proceeding papers* indexados na base de dados o qual é representado por $\frac{Art_i}{\sum_{j=1}^N Art_j} \sum_{k \in DN} PSJR_k$. O aumento dos periódicos ao longo dos anos poderá incluir também periódicos deste tipo e portanto aumentar o valor correspondente a esta parcela.

Especial cuidado deve ser dada à discussão apresentada relativamente a este aspeto, pois seria necessário conhecer em detalhe o cálculo dos diferentes parâmetros para inferir conclusões definitivas sobre a influência do aumento do número de periódicos ao longo dos anos no valor do *SJR*.

Complementaridade dos indicadores

A utilização de um único indicador para avaliar o impacto e o prestígio de periódicos não é suficiente devido à complexidade dos fatores que influenciam estes aspetos. Perante isto analisou-se se a informação fornecida pelo *JRI* poderá ser complementada pelo *SNIP* e *SJR* e vice-versa. As diferenças encontradas nas ordenações obtidas foram estudadas para um conjunto de periódicos. Na Figura 9 estão representadas as posições de um dado periódico na ordenação obtida utilizando o *JRI*, *SNIP* ou *SJR*. Foi utilizado o mesmo conjunto de periódicos a que se recorreu na análise do poder de previsão.

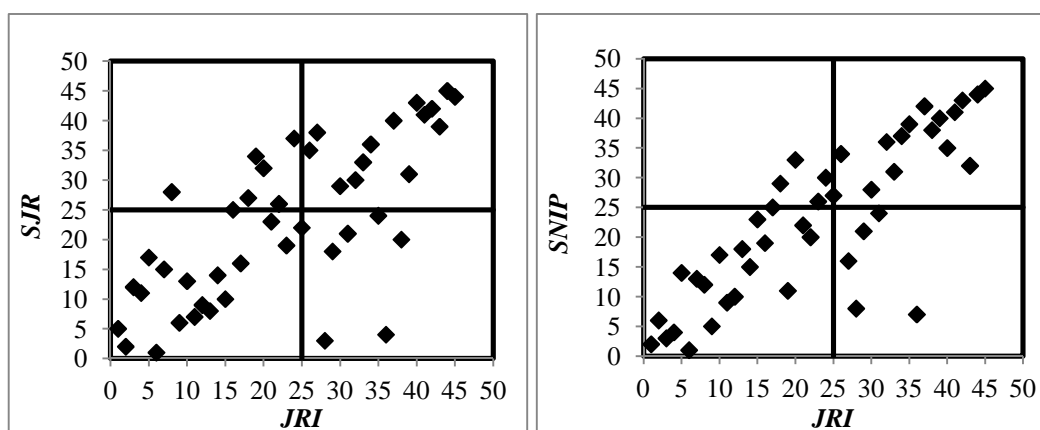


Figura 9. Diferenças observadas nas ordenações para o *JRI*, *SNIP* e *SJR*.

Consideráveis diferenças nas ordenações podem ser observadas na Figura 9. Em seguida são discutidos os vários aspetos considerados na definição do *JRI* e que podem justificar as vantagens deste indicador relativamente ao *SNIP* e *SJR*.

Tipologia dos documentos

Os periódicos que publicam apenas artigos de revisão ou uma elevada percentagem de artigos de revisão conseguirão atrair um elevado número de citações. No cálculo do *SNIP* e *SJR* o tipo de documento não é considerado. Portanto, para periódicos com uma elevada percentagem de artigos de revisão são esperados valores mais elevados de *SNIP* e *SJR* relativamente aos periódicos que não publicam artigos de revisão (Moed 2010; Gonzalez-Pereira, Guerrero-Bote e Moya-Anegon 2010). No *JRI* a média de citações por documento de um dado tipo é comparada com a média de citações por documento do mesmo tipo e que pertencem à mesma categoria. Assim é esperado que periódicos que publicam apenas artigos de revisão ou uma elevada percentagem de artigos de revisão ocupem uma melhor posição nas ordenações pelo *SNIP* e *SJR* do que nas ordenações pelo *JRI*. Os periódicos *Current Opinion In Colloid & Interface Science* e *International Reviews in Physical Chemistry* são dois casos que exemplificam o mencionado. O primeiro ocupa a 4ª e 7ª posição na ordenação pelo *SJR* e *SNIP* respetivamente, enquanto que na ordenação obtida pelo *JRI* ocupa a 36ª posição. O último ocupa a 3ª e 8ª posição na ordenação pelo *SJR* e *SNIP* respetivamente e a 28ª posição na ordenação do *JRI*.

Janela de citação

Na determinação do *SJR* a janela de citação varia entre 4 e 1 ano. Se um periódico publica documentos de uma área científica que recebe um número considerável de citações vários anos após publicação (mais de 3 anos) provavelmente ocupará uma posição baixa na ordenação final quando é utilizado o *SJR*. A janela de citação definida no *JRI* pretende corrigir para esta limitação e permite o acompanhamento do comportamento das citações ao longo do tempo de modo mais eficiente.

Normalização

O *SJR* tende a ser superior para categorias caracterizadas por uma comunidade científica que cita mais e portanto os periódicos classificados nestas categorias ocuparão uma melhor posição na ordenação final do que aqueles que pertencem a categorias onde se cita menos.

O estudo efetuado por Gonzalez-Pereira, Guerrero-Bote e Moya-Anegon 2010 mostra que os periódicos no topo 10% (ordenação decrescente pelo valor de *SJR*) na categoria *Genetics & Molecular Biology* têm valores de *SJR* muito superiores aos obtidos pelos periódicos classificados na categoria *Psychology*. Considerando a fórmula de cálculo deste indicador é possível analisar o efeito das diferentes culturas de citação.

$$PSJR_i = \underbrace{\frac{(1 - d - c)}{N}}_1 + \underbrace{e \frac{Art_i}{\sum_{j=1}^N Art_j}}_2 + d \underbrace{\left[\sum_{j=1}^N C_{ji} \frac{PSJR_j}{C_j} CF + \frac{Art_i}{\sum_{j=1}^N Art_j} \sum_{k \in DN} PSJR_k \right]}_3 \quad (40)$$

Para áreas caracterizadas por uma tendência maior para a citação o somatório $\sum_{j=1}^N C_{ji} \frac{PSJR}{C_j} CF$ irá ser superior ao obtido para as áreas que citam menos. Quanto maior o número de periódicos que citam um dado periódico maior o *SJR* obtido. Alguma normalização deveria ser introduzida neste ponto de modo a eliminar esta limitação. O *JRI* não sofre desta limitação pois o número total de documentos publicados numa dada categoria é utilizado na normalização, tornando-o independente da dimensão da categoria.

O *SNIP* determinado para um periódico pode ser muito afetado se o mesmo receber uma citação de um periódico com um elevado número de referências ativas. Para evidenciar esta limitação apresenta-se o seguinte exemplo.

Exemplo

Considere-se um periódico que publicou 30 documentos e que obteve em média 3 citações por documento (*RIP*) provenientes de 10 documentos citantes. A média de referências ativas por documento citante é de 4. O periódico que se encontra na posição mediana possui em média 3 referências ativas por documento. Para o periódico referido o *SNIP* é:

$$SNIP = \frac{3}{4/3} = 2,25 \quad (41)$$

Suponha-se agora que o periódico recebe duas citações adicionais provenientes de um documento citante com um elevado número de referências ativas, 60. O novo *RIP* do periódico será de 3,07 e a média de referências ativas passará a ser de 9,10. O novo valor para o *SNIP* será:

$$SNIP = \frac{3,07}{9,10/3} = 1,01 \quad (42)$$

O resultado obtido prova que a normalização adotada no *SNIP* não será a mais indicada, pois o facto de um periódico receber citações adicionais de documentos citantes com um elevado número de referências ativas, ao contrário do que seria esperado, não implica uma evolução positiva do valor do *SNIP*.

No caso do *JRI* médias aritméticas são utilizadas para determinar a média de citações por documento de um dado periódico e a média de citações por documento que caracteriza a(s) categoria(s) em que o periódico está classificado. Com o procedimento adotado a obtenção de uma citação adicional não implica a diminuição do *JRI*.

Até este ponto foram discutidos os aspetos que mostram as vantagens do *JRI* relativamente ao *SNIP* e *SJR*. O *SNIP* e *SJR* apresentam também as suas vantagens tornando-os indicadores que permitem complementar a informação fornecida pelo *JRI*.

O prestígio do periódico citante

As citações recebidas pelos documentos de um periódico podem ser provenientes de um periódico com baixo/elevado impacto na comunidade científica. No cálculo do *JRI* este aspeto não é levado em consideração, mas no *SJR* sim. Diferenças nas ordenações podem ser justificadas por este fator. É neste sentido que ambos os indicadores podem ser usados para fornecer mais informação acerca do impacto de um periódico na comunidade científica.

Normalização

No *SNIP* a normalização adotada é baseada no *citing-side* e portanto independente de um esquema de classificação dos periódicos em categorias. A utilização do *SNIP* em conjunto com o *JRI* poderá ser útil para um periódico que pertence a uma área muito especializada dentro de uma categoria. Embora se pretenda que as categorias sejam o mais homogêneas possível é difícil estabelecer uma fronteira.

A discussão dos resultados obtidos deve ser tida com muito cuidado pois seria necessário conhecer em detalhe os resultados dos vários parâmetros considerados no cálculo do *SNIP* e *SJR* para inferir se os vários aspetos discutidos justificam na realidade as diferenças observadas nas ordenações. Dada a indisponibilidade destes dados apenas podem ser apresentadas possíveis explicações.

Após o estudo apresentado para o *JRI* vários pontos fortes, do indicador, podem ser mencionados:

- a utilização da metodologia diacrónica na contabilização de citações assim como uma janela de citação longa e variável permite um acompanhamento mais eficiente do comportamento das citações em cada categoria. Este tipo de janela de citação é particularmente importante para periódicos que pertencem a categorias em que os documentos recebem um número considerável de citações vários anos após o ano de publicação;
- em cada categoria apenas são considerados os tipos de documentos que recebem 5% ou mais do total de citações obtidas por todos os documentos da categoria num período de 5 anos. Com isto pretende-se que a inclusão de um determinado tipo de documento seja apoiada pelo rendimento de cada tipo de documento relativamente às citações alcançadas e não por uma ideia preconcebida acerca da relevância das citações que é impossível generalizar;
- para periódicos classificados em mais do que uma categoria o desempenho do mesmo relativamente a cada categoria é avaliado e no final um valor médio é utilizado;
- o *JRI* é um indicador estável ao longo do tempo e com um poder de previsão razoável permitindo a um investigador uma seleção adequada do periódico para a publicação dos resultados de investigação.

A comparação do *JRI* com o *SNIP* e *SJR*, indicadores disponíveis, nas bases de dados *Scopus*, permitiu concluir que estes indicadores devem ser utilizados em conjunto quando se

pretende fornecer informação mais detalhada sobre o impacto alcançado por um periódico. O *JRI* revelou ser um indicador mais vantajoso do que o *SJR* e o *SNIP* nos seguintes aspetos:

- o tipo de documento não é considerado no *SNIP* e no *SJR* e portanto periódicos que publicam apenas artigos de revisão ou uma elevada percentagem deste tipo de documentos irão obter um elevado valor para o *SNIP* e *SJR*;
- no *SJR* a janela de citação utilizada privilegia os periódicos que publicam documentos que obtêm um número de citações considerável num pequeno espaço de tempo. A utilização de uma janela baseada na metodologia sincrónica torna os indicadores *SJR* e *SNIP* mais vulneráveis a oscilações no comportamento das citações;
- o *SJR* tende a ser superior para os periódicos que pertencem a áreas científicas onde se cita mais. No entanto, algumas alterações foram introduzidas muito recentemente no *SJR* no sentido de permitir a comparação entre periódicos de diferentes áreas (Scopus 2013a);
- na normalização aplicada no *SNIP* a obtenção de uma citação adicional proveniente de um documento citante com elevado número de referências ativas poderá contribuir para a diminuição do *SNIP*. A obtenção de citações deve permitir a evolução do indicador no sentido positivo, pois acumular citações é positivo. No *JRI* isto não se verifica pois a obtenção de uma citação adicional não irá diminuir o valor do *JRI*. No entanto alterações recentes da metodologia de cálculo do *SNIP* permite corrigir para esta limitação (Waltman, van Eck, van Leeuwen *et al.* 2013).
- o *JRI* é mais estável do que o *SJR* e *SNIP* permitindo uma seleção mais segura ao investigador do periódico onde pretende publicar os seus resultados.

Apesar dos pontos fortes e das vantagens relativamente ao *SNIP* e *SJR*, o *JRI* também apresenta limitações:

- o tipo de normalização adotado, *cited-side*, é dependente do esquema de classificação adotado para os periódicos, neste caso as categorias definidas pelo *WoS*;
- as longas janelas de citação utilizadas no *JRI* não permitem fornecer informação acerca do impacto de um periódico num pequeno espaço de tempo. Dependendo da categoria a que pertence serão necessários vários anos de indexação no *WoS* para se obter resultados. Isto constitui uma limitação principalmente para periódicos que aparecem e desaparecem em pequenos períodos de tempo;
- o cálculo deste indicador é bastante complexo devido à utilização de janelas de citação variáveis tornando difícil a sua implementação em bases de dados como o *Scopus* e *WoS*.

Atendendo ao facto mencionado no último ponto e por ser necessário uma grande quantidade de dados que teriam que ser retirados do *WoS*, este indicador não foi utilizado posteriormente na definição do modelo pretendido. No entanto o indicador é aqui apresentado pois considera-se que apresenta um conjunto de propriedades importantes quando utilizado para caracterizar o desempenho científico de um periódico e que pode complementar a informação disponibilizada por outros indicadores destinados para o mesmo efeito.

3.3.2 O índice h_{nf}

O h_{nf} é um indicador que permite caracterizar o desempenho científico de um investigador. É baseado no conceito do índice h corrigido para algumas das suas limitações e que preserva outras propriedades do índice que são desejáveis. Em seguida é efetuada uma análise detalhada deste indicador.

Aplicação do h_{nf} a um conjunto de indivíduos

O h_{nf} foi determinado para um conjunto de investigadores holandeses da área de *Pharmacology & Pharmacy (PP)* e *Mathematics, Applied (MA)*. Uma descrição detalhada da amostra encontra-se na secção 3.1.

Uma vez que o h_{nf} é baseado no conceito do índice h os resultados obtidos com a utilização do novo indicador são comparados com os obtidos pelo índice h .

O índice h foi inicialmente desenvolvido para ser aplicado a investigadores e rapidamente se difundiu na comunidade científica pelas propriedades apresentadas. Vários estudos foram efetuados com o objetivo de estudar a relação do índice h com o número de documentos (Costas e Bordons 2007; Molinari e Molinari 2008; Schubert e Glanzel 2007). A principal conclusão destes estudos é que o índice h não é independente do número de documentos. Quanto maior o número de documentos maior o valor esperado para este índice. Ao introduzir o fracionamento no cálculo do índice h_{nf} espera-se conseguir uma comparação mais eficiente de investigadores que trabalham em áreas científicas com culturas de publicação muito diferentes (por exemplo matemática e química). Para uma melhor compreensão dos efeitos introduzidos pelo fracionamento dos documentos nas Tabelas 17 e 18 é apresentado o valor do índice h e do h_f . No h_f cada documento é fracionado pelo número de autores, as citações não são alteradas e o procedimento para o seu cálculo é semelhante a aplicado ao índice h .

O fracionamento dos documentos resulta em h_f inferiores ao índice h o que está de acordo com o esperado dada a dependência do índice em relação ao número de documentos publicados. Para os investigadores considerados observou-se que na *Pharmacology & Pharmacy* existem valores de h_f que são inferiores entre 40% e 70% ao valor obtido para o índice h . Na *Mathematics, Applied* o investigador B possui o mesmo valor para o h_f e índice h pois é um investigador que colabora pouco. Este investigador possui em média 1,7 autores por documento. Para esta categoria os valores de h_f e do índice h estão mais próximos quando comparados com as diferenças observadas na *Pharmacology & Pharmacy*. A diferença máxima foi observada para o investigador K que possui um h_f inferior em 46% em relação ao índice h . Este comportamento era esperado pois a categoria *Pharmacology & Pharmacy* é caracterizada por um maior número de colaborações e portanto o fracionamento tem um grande efeito. Para o conjunto de investigadores aqui considerados a média de autores por documento encontrada foi de 6,0 e 2,4 para a *Pharmacology & Pharmacy* e *Mathematics, Applied* respetivamente.

No entanto, o importante aqui é a comparação das ordenações obtidas quando se utiliza o índice h e o h_f pois é isto que permite verificar como se comporta o índice h_f para indivíduos de diferentes áreas. Na *Pharmacology & Pharmacy* existem três investigadores que ocupam a mesma posição em ambas as ordenações enquanto que em *Mathematics, Applied* nenhum dos investigadores possui a mesma posição. Os investigadores com um elevado número de autores por documento foram na sua maioria colocados numa posição inferior na ordenação pelo h_f relativamente à ordenação com o índice h . As posições ocupadas pelos investigadores N e P de *Pharmacology & Pharmacy* e E, N e P de *Mathematics, Applied* na ordenação pelo índice h_f não são justificadas pelo fracionamento pois trata-se de investigadores com um baixo número de autores por documento. O baixo número de citações dos documentos poderá ser a justificação para os resultados observados, no entanto isto apenas será confirmado após a normalização das citações, isto é, determinando o h_{nf} para cada investigador.

Tabela 17. Valores determinados para o índice h e h_f na *PP*.

Investigador	h	Ordenação (h)	h_f	Ordenação (h_f)
A	29	1	11,84	1
D	22	2	9,08	2
H	18	3	6,00	7
E	17	4	7,39	3
J	16	5	5,72	10
L	16	5	5,92	9
G	15	6	6,64	6
M	15	6	5,59	11
C	13	7	6,97	4
I	12	8	6,83	5
Q	12	8	4,95	13
K	11	9	5,93	8
O	11	9	4,86	14
B	10	10	5,58	12
N	10	10	4,70	16
P	9	11	4,49	17
F	8	12	4,83	15

Tabela 18. Valores determinados para o índice h e h_f na MA.

Investigador	h	Ordenação (h)	h_f	Ordenação (h_f)
G	7	1	4,26	3
D	6	2	4,53	1
I	6	2	3,42	6
A	5	3	4,50	2
C	5	3	4,50	2
F	5	3	3,83	4
H	5	3	3,50	5
J	5	3	3,17	7
K	5	3	2,67	10
E	4	4	2,83	9
M	4	4	2,33	11
Q	4	4	2,25	12
B	3	5	3,00	8
L	3	5	2,83	9
N	3	5	2,17	13
O	3	5	2,08	14
P	3	5	1,83	15

Uma vez avaliado o efeito do fracionamento é possível agora analisar as alterações introduzidas pela normalização.

Os valores obtidos para o h_{nf} e o índice h são apresentados nas Tabelas 19 e 20. A normalização *cited-side* e a metodologia diacrónica foram adotadas no processo de normalização do h_{nf} .

Tabela 19. Valores obtidos para o índice h e h_{nf} na PP.

Investigador	h	Ordenação (h)	h_{nf}	Ordenação (h_{nf})
A	29	1	10,73	1
D	22	2	8,08	2
H	18	3	5,60	11
E	17	4	6,46	6
J	16	5	5,68	10
L	16	5	6,16	9
G	15	6	6,56	5
M	15	6	5,53	12
C	13	7	7,34	3
I	12	8	7,08	4
Q	12	8	4,61	14
K	11	9	6,24	7
O	11	9	4,65	13
B	10	10	6,17	8
N	10	10	4,25	17
P	9	11	4,33	16
F	8	12	4,53	15

Tabela 20. Valores obtidos para o índice h e h_{nf} na MA.

Investigador	h	Ordenação (h)	h_{nf}	Ordenação (h_{nf})
G	7	1	4,84	7
D	6	2	5,03	5
I	6	2	4,42	9
A	5	3	6,83	1
C	5	3	5,83	2
F	5	3	5,33	4
H	5	3	5,58	3
J	5	3	3,67	11
K	5	3	3,00	13
E	4	4	4,67	8
M	4	4	3,00	13
Q	4	4	3,08	12
B	3	5	5,00	6
L	3	5	4,08	10
N	3	5	2,83	14
O	3	5	2,83	14
P	3	5	2,17	15

Para a categoria *Mathematics, Applied* o h_{nf} é da mesma ordem de grandeza do índice h ao contrário do que acontece para a *Pharmacology & Pharmacy*.

Na *Pharmacology & Pharmacy* os valores de h_{nf} são inferiores aos do índice h , enquanto que na *Mathematics, Applied* existem investigadores em que o valor obtido é superior. A *Pharmacology & Pharmacy* é caracterizada por uma elevada média de citações por documento quando comparada com a *Mathematics, Applied*. Na recolha dos dados para a normalização verificou-se que os artigos de *Pharmacology & Pharmacy* publicados em 2000 cujas citações foram contabilizadas no período 2000-2007 obtiveram em média 14,7 citações enquanto que os artigos de *Mathematics, Applied* publicados no mesmo ano e com as citações recolhidas no mesmo período obtiveram em média 6,5 citações. Com o processo de normalização adotado existe uma depreciação das citações obtidas na *Pharmacology & Pharmacy* enquanto que na *Mathematics, Applied* o comportamento é o inverso. Para os artigos de todas as categorias publicados em 2000, com as citações contabilizadas em 2000-2007 o número médio de citações por documento encontrado foi de 15,9.

A introdução da normalização permitiu acentuar mais as diferenças nas ordenações obtidas quando cada um dos índices é utilizado. Por exemplo o investigador H de *Pharmacology & Pharmacy* que com o fracionamento ocupava a posição sete na ordenação pelo h_f (e posição três na ordenação pelo h) ocupa agora a posição onze na ordenação pelo h_{nf} . Este investigador publica em outras categorias que são caracterizadas por um número médio de citações por documento mais elevado do que em *Pharmacology & Pharmacy* e portanto a depreciação das citações obtidas é ainda superior. A normalização também garantiu uma melhor posição a alguns investigadores relativamente à ordenação com o índice h e à ordenação com o h_f .

A utilização do h_{nf} permite diferenciar melhor os investigadores. De facto os valores obtidos para o h_{nf} sugerem que existem investigadores com documentos cujo desempenho em

relação ao número de citações é diferente enquanto que com o índice h isto não é visível (exemplo: investigador C, F, H, J e K de *Mathematics, Applied*).

Na Tabela 21 são apresentados alguns índices que permitem caracterizar a amostra de investigadores utilizada relativamente a alguns indicadores.

Tabela 21. Média, desvio padrão e coeficiente de variação para diferentes indicadores.

Índice	Estatísticas	<i>Mathematics, Applied</i>	<i>Pharmacology & Pharmacy</i>
TD	Média	20,88	118,00
	Desvio padrão	6,40	50,36
	Coeficiente de variação (%)	31	43
NDF	Média	10,19	23,31
	Desvio padrão	3,23	9,00
	Coeficiente de variação (%)	32	39
Índice h	Média	4,47	14,35
	Desvio padrão	1,23	5,26
	Coeficiente de variação (%)	28	37
Índice h_{nf}	Média	4,25	6,12
	Desvio padrão	1,30	1,63
	Coeficiente de variação (%)	31	27

Para os investigadores de *Pharmacology & Pharmacy* o índice h é em média superior ao observado para a categoria *Mathematics, Applied*. O desvio padrão e o coeficiente de variação associados são mais elevados também. Este resultado sugere que os investigadores de *Pharmacology & Pharmacy* constituem uma amostra mais heterogénea. Dois fatores poderão explicar o observado:

- o número de documentos publicados (TD). Para a amostra considerada de investigadores na *Pharmacology & Pharmacy* os desvios em relação à média atingem, em média, 43% do valor desta. A contagem dos documentos após a aplicação do fracionamento (NDF) resulta num coeficiente de dispersão ligeiramente inferior;
- a distribuição dos documentos por um conjunto de categorias com culturas de citação muito diferentes.

A aplicação da normalização e o fracionamento resulta em valores de h_{nf} que mostram que os investigadores de *Pharmacology & Pharmacy* estão mais próximos entre si relativamente ao seu desempenho. O desvio padrão obtido na *Pharmacology & Pharmacy* para o h_{nf} é agora muito próximo do obtido para o h_{nf} em *Mathematics, Applied*. Estes resultados devem ser tidos com muito cuidado pois apenas se formularam conclusões sobre uma fração dos documentos publicados, isto é, pelos documentos que definem o índice h e h_{nf} .

Após uma análise das alterações introduzidas pela utilização de um indicador que considera o número de autores por documento e que normaliza as citações é apresentada na Tabela 22 a ordenação dos 34 investigadores por ordem decrescente do valor de h_{nf} . Outros indicadores são também apresentados com o objetivo de auxiliar na discussão dos resultados.

Tabela 22. Ordenação decrescente dos investigadores utilizando os valores de h_{nf} .

Investigadores	SC	TD	APD	CPD	NDF	h	TD h -core (%)	h_{nf}	TD h_{nf} -core (%)
A	PP	294	7,42	11,86	50,73	29	10	10,73	23
D	PP	134	3,94	10,27	23,51	22	16	8,08	37
C	PP	111	5,57	5,65	25,72	13	12	7,34	27
I	PP	122	5,19	5,23	27,17	12	10	7,08	30
A	MA	14	1,64	4,79	10,50	5	36	6,83	57
G	PP	147	5,55	4,54	28,52	15	10	6,56	21
E	PP	85	6,64	11,15	16,01	17	20	6,46	44
K	PP	75	6,35	5,28	16,18	11	15	6,24	37
B	PP	100	6,56	3,95	34,26	10	10	6,17	20
L	PP	146	4,11	4,52	25,39	16	11	6,16	24
C	MA	29	2,59	2,52	12,75	5	17	5,83	41
J	PP	104	6,31	6,5	17,21	16	15	5,68	33
H	PP	94	6,70	10,74	15,54	18	19	5,60	38
H	MA	30	2,40	2,87	15,33	5	17	5,58	47
M	PP	82	6,70	8	13,06	15	18	5,53	43
F	MA	28	2,50	2,39	12,08	5	18	5,33	39
D	MA	18	2,17	10,17	9,70	6	33	5,03	56
B	MA	15	1,67	2,73	11,53	3	20	5,00	40
G	MA	22	3,05	5,36	8,84	7	32	4,84	50
E	MA	32	1,97	1,53	18,83	4	13	4,67	25
O	PP	124	7,18	3,45	21,47	11	9	4,65	22
Q	PP	108	7,14	5,06	18,11	12	11	4,61	27
F	PP	81	7,04	3,37	23,77	8	10	4,53	22
I	MA	22	3,27	7,32	7,79	6	27	4,42	55
P	PP	108	5,31	10,89	22,09	9	8	4,33	19
N	PP	91	5,68	3,73	17,52	10	11	4,25	23
L	MA	18	2,22	1,78	9,17	3	17	4,08	44
J	MA	19	3,11	5,21	6,42	5	26	3,67	53
Q	MA	30	3,00	1,63	10,83	4	13	3,08	30
M	MA	16	2,56	2,38	6,67	4	25	3,00	44
K	MA	16	2,50	3,06	7,17	5	31	3,00	44
N	MA	15	1,93	1,73	8,67	3	20	2,83	33
O	MA	17	3,00	2,12	6,81	3	18	2,83	35
P	MA	14	1,64	0,79	10,17	3	21	2,17	29

Não é surpreendente que os investigadores de *Pharmacology & Pharmacy* ocupem os lugares de topo. Na realidade as alterações introduzidas compensam para as diferentes culturas de citações de cada categoria através do processo de normalização, mas o

fracionamento considerado não é suficiente para corrigir as diferentes culturas de publicação. A intensa colaboração (medida pela número de autores por documento) em *Pharmacology & Pharmacy* relativamente à *Mathematics, Applied* justifica em parte o elevado número de documentos publicados. A contagem fracionada corrige para este fator, mas no final os investigadores de *Pharmacology & Pharmacy* possuem ainda um maior número de documentos. Assim como o índice h também o h_{nf} não é independente do número de documentos publicados e portanto a ordenação final é influenciada em certa medida por este fator. Note-se que os investigadores colocados nas posições de topo encontram-se entre os que têm um maior número de documentos publicados. Isto sugere que uma correção adicional para o número de documentos deveria ainda ser introduzida de modo a permitir uma comparação completa de investigadores que publicam em diferentes categorias. A quantidade de documentos publicados é também influenciada por outros aspetos tais como o tempo dedicado à investigação, o financiamento disponível entre outros. Indicadores baseados nestes aspetos são difíceis de obter, mas quando disponíveis permitem uma comparação mais justa dos investigadores.

A percentagem de documentos que definem o h_{nf} é superior à obtida para o índice h . Quanto maior a percentagem de documentos considerada melhor a descrição do desempenho de cada indivíduo quando indicadores deste tipo são utilizados.

Poder de previsão do h_{nf}

Indicadores com algum poder de previsão são desejados se o objetivo passa também por inferir sobre qual será o desempenho científico de um investigador no futuro. O poder de previsão do h_{nf} é aqui avaliado e comparado com o de outros indicadores que são muitas vezes utilizados para descrever o desempenho científico de indivíduos.

Na Tabela 23 o número total de documentos (TD), o número total de citações (TC), o índice h e o índice h_{nf} estão representados para dois períodos, 1996-2003 e 2000-2007. Os indicadores referidos foram determinados para um conjunto de investigadores da categoria *Pharmacology & Pharmacy*.

Tabela 23. Valores de h_{nf} , h , TC e TD em 1996-2003 e 2000-2007 para a *PP*.

Investigadores	1996-2003				2000-2007			
	h_{nf}	h	TC	TD	h_{nf}	h	TC	TD
H	8,36	21	1627	142	5,60	18	1010	94
F	7,96	13	659	180	4,17	8	273	81
I	7,64	13	691	112	7,08	12	638	122
A	7,37	23	1639	173	10,73	29	3486	294
D	6,80	17	916	135	5,60	22	1376	134
E	6,42	15	750	89	6,46	17	948	85
Q	5,13	11	514	122	4,61	12	546	108
G	3,89	10	269	57	6,56	15	667	146
M	3,41	8	255	54	5,53	15	656	82
N	3,16	6	98	44	4,25	10	339	91

A análise deve ser efetuada com precaução pois os critérios de seleção dos investigadores poderão conduzir a um conjunto heterogéneo. Alguns investigadores poderão ser ainda relativamente novos e estarem numa fase de ascensão na sua carreira científica, outros desenvolveram já todo o seu potencial e estarão no auge da carreira e outros existirão já numa fase final da carreira profissional.

O índice h e o número total de citações são os indicadores em que um menor número de investigadores apresentam valores inferiores para o período 2000-2007. A obtenção de prestígio ao longo da carreira profissional do investigador poderá explicar o comportamento observado. Inicialmente o trabalho científico desenvolvido por um investigador poderá não ter um impacto elevado na comunidade científica e algum tempo será necessário até que este seja reconhecido na comunidade. Perante isto é esperado que ao longo do tempo o impacto na comunidade científica seja crescente e portanto o valor do índice h e do número total de citações seja mais elevado.

No conjunto de investigadores utilizados o investigador I é aquele que apresenta um comportamento mais destacado dos restantes. Aparece nas posições de topo quando o h_{nf} é utilizado, ao contrário do que acontece nos restantes indicadores. A maioria dos periódicos onde este investigador publicou estão classificados apenas na categoria *Pharmacology & Pharmacy*, enquanto que os restantes investigadores publicaram em periódicos, na sua maioria, classificados também em outras categorias e que são caracterizadas por uma média de citações por documento superior. Com a normalização aplicada no h_{nf} as citações obtidas pelos documentos destes investigadores são depreciadas colocando-os numa posição inferior em relação ao investigador I.

A ordenação decrescente dos investigadores para cada indicador no período 1996-2003 permite identificar a posição relativa de um investigador em relação aos restantes. A adoção do mesmo procedimento para o período 2000-2007 permite a comparação de resultados e inferir sobre o poder de previsão de cada um dos indicadores. No total foram contabilizados 45 pares. Dos 45 pares em cerca de 58% a posição dos 2 investigadores de cada par permanece igual quando o h_{nf} é utilizado. Este valor é superior quando o TC e o índice h são utilizados, 80% em ambos os casos. O menor valor verifica-se para o TD , 56%. No entanto os resultados obtidos merecem ser complementados com um breve comentário. Para o investigador F é possível verificar que o TD diminui consideravelmente, cerca de 55%, para o período 2000-2007. Isto poderá indicar que o investigador em causa se “dirige” para uma fase final da sua carreira. Tratando-se de uma situação deste género é aceitável que os valores dos indicadores considerados sejam inferiores para o período mais recente. Todos os indicadores mencionados são dependentes da dimensão, isto é, do número de documentos publicados. Portanto se o TD diminui é esperada também uma diminuição dos restantes indicadores. Se na mesma análise o investigador F é excluído o TC e o índice h continuam a ser os indicadores com maior poder de previsão (86%). Para o h_{nf} e o TD as percentagens aumentam para 75% e 69% respetivamente. Estes resultados evidenciam que o poder de previsão do índice h e do h_{nf} não são tão diferentes como sugerido pela análise inicial (análise com os 10 investigadores). Portanto o facto de o h_{nf} permitir a comparação de indivíduos que trabalham em diferentes áreas torna-o mais adequado do que o h para descrever o desempenho científico.

Comparação do h_{nf} com variantes do índice h

Desde o surgimento do índice h , em 2005, que vários estudos têm sido feitos ao longo dos anos com o objetivo de identificar as vantagens e desvantagens do indicador. Outros estudos sugerem novos indicadores baseados no conceito do índice h , mas que foram alterados no sentido de melhorar a informação fornecida.

Dos novos indicadores sugeridos alguns foram selecionados e determinados para o conjunto de investigadores aqui considerados. Com isto pretende-se comparar as ordenações obtidas utilizando os diferentes indicadores e mostrar as vantagens da utilização do h_{nf} relativamente a outros já existentes.

Os indicadores foram determinados utilizando a produção científica indexada no *WoS* no período 2000-2007 da amostra de investigadores aqui considerada. Para cada índice é apresentada a fórmula utilizada no seu cálculo e discutidas as alterações observadas nas ordenações apresentadas na Tabela 24 utilizando alguns investigadores como exemplo.

Tabela 24. Ordenação decrescente dos investigadores para cada indicador.

Investigadores	Categoria	Ordenação							
		h_{nf}	A	R	AR	g	hg	h_m	h_l
A	PP	1	1	1	1	1	1	1	2
D	PP	2	3	2	2	2	2	2	6
C	PP	3	9	9	11	9	9	7	3
I	PP	4	15	11	12	11	11	8	16
A	MA	5	22	22	21	22	22	22	1
G	PP	6	5	5	5	6	8	9	7
E	PP	7	4	4	3	4	4	4	5
K	PP	8	19	15	16	16	14	12	15
B	PP	9	13	13	8	13	13	14	11
L	PP	10	11	8	9	8	7	10	13
C	MA	11	31	27	31	30	26	24	18
J	PP	12	10	7	7	7	5	6	12
H	PP	13	2	3	4	3	3	3	8
H	MA	14	24	23	24	23	23	23	22
M	PP	15	6	6	6	5	6	5	21
F	MA	16	27	24	25	24	24	25	10
D	MA	17	8	17	19	17	18	13	9
B	MA	18	26	31	30	31	31	27	4
G	MA	19	21	20	20	20	20	20	20
E	MA	20	29	26	26	26	27	31	14
O	PP	21	14	12	13	12	12	15	27
Q	PP	22	7	10	10	10	10	11	28
F	PP	23	18	18	17	18	17	18	19
I	MA	24	12	19	18	19	19	17	23
P	PP	25	17	16	15	15	16	19	26
N	PP	26	16	14	14	14	15	16	24
L	MA	27	33	33	33	33	33	32	32
J	MA	28	20	21	22	21	21	21	25
Q	MA	29	30	29	27	27	28	29	33
M	MA	30	25	30	29	29	29	28	34
K	MA	31	28	25	23	25	25	26	17
N	MA	32	32	32	32	32	32	33	30
O	MA	33	23	28	28	28	30	30	31
P	MA	34	34	34	34	34	34	34	29

O índice A representa a média de citações por documento dos documentos que definem o índice h . A seguinte fórmula é usada no cálculo do índice A (Jin, Liang, Rousseau *et al.* 2007):

$$A = \frac{1}{h} \sum_{j=1}^h C_j \quad (43)$$

onde C_j é o número de citações, obtidas num determinado período, pelo documento j .

O índice R representa a raiz quadrada da soma das citações obtidas pelos documentos que definem o índice h (Jin, Liang, Rousseau *et al.* 2007):

$$R = \sqrt{\sum_{j=1}^h C_j} \quad (44)$$

O índice AR é calculado de modo semelhante ao R , mas a idade de cada documento é considerada (Jin, Liang, Rousseau *et al.* 2007):

$$AR = \sqrt{\sum_{j=1}^h \frac{C_j}{a_j}} \quad (45)$$

onde a_j representa o número de anos após publicação do documento j .

No cálculo do índice g todos os documentos são ordenados por ordem decrescente do número de citações. As citações acumuladas são contabilizadas. Um investigador tem um índice g se o topo g documentos somam no mínimo um total de g^2 citações (Egghe 2006). De modo a elucidar o cálculo do índice um exemplo é mostrado na Tabela 25. Para este exemplo o índice g será de 5.

Tabela 25. Cálculo do índice g .

Ordenação (r)	TC	TC acumuladas	Ordenação (r ²)
1	11	11	1
2	10	21	4
3	5	26	9
4	3	29	16
5	3	32	25
6	2	34	36
...

O índice h_m ou o “*mock h index*” é obtido do seguinte modo (Prathap 2010):

$$h_m = \left(\frac{(\sum_{j=1}^P C_j)^2}{P} \right)^{\frac{1}{3}} \quad (46)$$

onde P é o número total de documentos publicados por um investigador e C_j o número total de citações de cada documento.

O índice hg é a média geométrica dos índices h e g determinados para um investigador (Alonso, Cabrerizo, Herrera-Viedma *et al.* 2010).

$$hg = \sqrt{h \times g} \quad (47)$$

Utilizando a média geométrica o valor de hg vai ser mais próximo do valor de h e portanto a influência dos documentos com um elevado número de citações que afetam o índice g irá ser atenuada.

Como é possível constatar pela descrição apresentada os índices A , R e AR contabilizam o número de citações obtidas pelos documentos que definem o índice h . O índice g , h_m e hg são índices que pretendem também contabilizar citações. No caso do h_m são consideradas as citações de todos os documentos.

As ordenações obtidas para estes indicadores (Tabela 24) mostram nas posições de topo e intermédias essencialmente investigadores da categoria *Pharmacology & Pharmacy*. Para estes índices a existência de documentos com um elevado número de citações influencia consideravelmente os resultados. Como exemplo tem-se os investigadores D e I de *Mathematics, Applied* e Q de *Pharmacology & Pharmacy* que apresentam posições muito diferentes quando a ordenação é efetuada pelo índice h_{nf} e índice A . Estes investigadores possuem dois documentos (investigadores D e I) e um documento (investigador Q), no conjunto de documentos que definem o índice h , com um elevado número de citações. A aplicação da raiz quadrada atenua o efeito dos documentos com muitas citações, mas não é suficiente. Os documentos muito citados dos investigadores I e Q mencionados anteriormente dominam o índice R , embora o efeito seja inferior. O índice AR atenua ainda mais o efeito destes documentos dividindo pela idade. No entanto a presença de documentos com um elevado número de citações não é a única explicação para as diferenças observadas nas ordenações. A ausência de normalização de citações e do fracionamento também explicam o comportamento observado. A normalização aplicada no h_{nf} permite a valorização das citações dos investigadores de áreas caracterizadas por um baixo número de citações por documento e o fracionamento diminui a discrepância entre o número de documentos publicados (embora por si só não seja suficiente) para as duas categorias. As diferenças observadas para os investigadores I de *Pharmacology & Pharmacy*, A e C de *Mathematics, Applied* são justificadas pela normalização e fracionamento.

O índice h_I considera no seu cálculo o número total de autores dos documentos que definem o índice h (Batista, Campiteli, Kinouchi *et al.* 2006).

$$h_I = \frac{h}{N_a} \quad (48)$$

onde N_a é o número médio de autores por documento do conjunto de documentos que definem o índice h .

Este índice é fortemente afetado pela existência de documentos com um elevado número de autores. No conjunto de investigadores aqui considerados não foi detetado nenhum caso nesta situação.

As ordenações obtidas quando os índices h_{nf} e h_I são utilizados mostram que a maioria dos investigadores de *Pharmacology & Pharmacy*, ocupam uma posição inferior à observada na ordenação pelo índice h_{nf} . Para os investigadores de *Mathematics, Applied* o comportamento oposto é observado. O facto de os autores de *Mathematics, Applied* possuírem em média menos autores por documento relativamente à *Pharmacology & Pharmacy* explica as diferenças observadas. O investigador A de *Mathematics, Applied* ocupa a primeira posição quando a ordenação é efetuada pelo índice h_I e o investigador B a posição quatro. Estes são investigadores com um pequeno número de autores por documento (1,2 para o investigador A e 1 para o investigador B). Apesar destes investigadores possuírem um baixo índice h quando

comparados com os obtidos pelos investigadores de *Pharmacology & Pharmacy*, o baixo número de autores por documento é suficiente para os colocar nas posições de topo.

No índice h_{nf} a divisão pelo número de autores documento a documento parece um procedimento mais correto. Documentos com elevado número de autores terão uma pequena contribuição na determinação do h_{nf} . Ao utilizar o fracionamento sugerido no h_I a influência destes documentos sobre o valor final do h_I é elevada e poderá prejudicar fortemente investigadores com um desempenho razoável, mas cuja cultura de publicação da área científica de trabalho é caracterizada por uma intensa colaboração.

A ordenação obtida com a utilização do h_{nf} mostra os investigadores B e D de *Mathematics, Applied* nas posições intermédias. Estes investigadores possuem um pequeno número de autores por documento, mas a normalização utilizada no h_{nf} sugere que os documentos destes autores receberam menos citações do que os documentos publicados pelos investigadores nas posições de topo no h_{nf} . Mais uma vez é evidenciada a necessidade da normalização das citações se o objetivo é comparar indivíduos com diferentes culturas de publicação e citação.

Os resultados apresentados até este ponto sugerem que o índice h_{nf} constitui uma boa alternativa ao índice h . Este mantém as propriedades atrativas do índice h :

- é um índice que fornece informação acerca da quantidade (número de documentos) e do impacto (número mínimo de citações dos documentos, após normalização, que definem o h_{nf});
- o índice h_{nf} não é afetado pelos documentos com elevado número de citações nem pelo conjunto de documentos com zero citações. Considere-se o caso de uma instituição de investigação de elevado prestígio que tem estado a desenvolver uma metodologia inovadora. Um investigador A de uma outra instituição deu o seu contributo no desenvolvimento da metodologia. A publicação resultante do trabalho poderá causar grande impacto na comunidade científica (medido pelo número de citações) no período logo após a sua publicação e ao longo dos vários anos. Esta publicação com muitas citações poderá ser uma observação atípica no total de citações contabilizadas e na média de citações por documento de um investigador. Se o h_{nf} é calculado de modo semelhante ao índice h esta publicação irá ter a mesma contribuição que terá a publicação com o segundo valor de citações mais elevado. No outro extremo estão os documentos a que van Raan 2004 chamou de *sleeping beauties* que irão puxar para “baixo” o número médio de citações por documento se este índice é utilizado, mas no caso do h_{nf} e h isto não se verifica. Van Raan usa o termo *sleeping beauties* para denominar os documentos que não obtêm citações num futuro imediato, mas apenas vários anos após o ano de publicação;
- permite prever o comportamento de um investigador no futuro com base na informação disponível no presente. Para o pequeno conjunto de investigadores avaliados foi possível prever corretamente a posição relativa dos indivíduos de 75% dos pares de investigadores disponíveis. Um pouco superior foi o valor determinado para o índice h , 86%.

O índice h_{nf} apresenta também outras propriedades:

- com o processo de normalização descrito o impacto de cada documento (medido pelo número de citações obtidas) é comparado com o impacto obtido pelos documentos do mesmo tipo, publicados no mesmo ano e em periódicos pertencentes à(s) mesma(s) categoria(s). Isto permite a comparação de indivíduos que trabalham em áreas científicas com diferentes culturas de citação;
- a extensão da lista de autores de cada documento é considerada. O fracionamento de cada documento pelo respetivo número de autores permite corrigir, em parte, para as diferentes culturas de publicação;
- a percentagem do total de documentos que definem o h_{nf} é superior à percentagem que define o índice h . Uma das desvantagens apontadas ao índice h é a incapacidade para descrever toda a produção científica de um investigador. No conjunto de investigadores aqui considerado a percentagem do total de documentos, publicados entre 2000 e 2007, que o índice h representa varia entre 9% e 36%. Para o índice h_{nf} os documentos que o definem representam entre 20% e 57% do total de documentos dos investigadores. Quanto maior esta percentagem mais informação é fornecida acerca do desempenho científico do indivíduo.

Apesar das propriedades atrativas apresentadas pelo índice h_{nf} este ainda apresenta algumas limitações que não permitem uma justa comparação entre indivíduos que pertencem a áreas científicas com diferentes culturas de publicação. O fracionamento tem em consideração as diferenças que existem relativamente à colaboração, mas não é suficiente para permitir uma completa comparação de investigadores que trabalham em áreas científicas caracterizadas por grandes diferenças relativamente à quantidade de documentos publicados. Por outro lado, existe o tipo de normalização adotado, *cited-side*, que é dependente do esquema de classificação adotado para os periódicos, neste caso as categorias definidas pelo WoS.

3.3.3 Impacto normalizado para investigadores (*Normalized Indicators for Researchers, NIR*)

O *NIR* é um indicador que procura dar informação acerca do impacto médio do conjunto de documentos publicados por um investigador. A média de citações por documento é utilizada muitas vezes para descrever o impacto médio dos documentos publicados. No entanto a utilização deste indicador sem se proceder inicialmente a uma normalização não o torna adequado para a comparação de indivíduos, instituições ou outro corpo de investigação com atividades em diferentes áreas. O *NIR* permite fazer esta comparação dado tratar-se de um indicador normalizado. Neste indicador foi utilizada a normalização *cited-side* e a metodologia diacrónica.

As várias propriedades do *NIR* são apresentadas em seguida. Uma comparação com o indicador média de citações por documento (*CPD*) é providenciada com o objetivo de evidenciar as vantagens do indicador.

Os resultados obtidos (Tabela 26 e 27) mostram alterações nas ordenações quando o indicador *NIR* é utilizado nas duas categorias. As alterações são mais pronunciadas na *Mathematics, Applied*. Para esta categoria apenas 18% dos investigadores não sofrem qualquer alteração na posição que detinham na ordenação obtida quando o indicador *CPD* é utilizado, enquanto que na categoria *Pharmacology & Pharmacy* cerca de 41% dos investigadores permanecem na mesma posição. De facto a normalização contribui para uma diminuição da dispersão observada. Quando o *NIR* é utilizado o coeficiente de variação da *Pharmacology & Pharmacy* passa de 48% (observado para *CPD*) para 41% e na *Mathematics, Applied* de 71% (observado para *CPD*) para 54%. Isto sugere que os indivíduos da categoria *Mathematics, Applied*, aqui apresentados, são mais heterogéneos em relação ao impacto da sua produção do que os indivíduos de *Pharmacology & Pharmacy*.

Tabela 26. Valores obtidos para o *CPD* e *NIR* na *PP*.

Investigador	CPD	Ordenação (CPD)	NIR	Ordenação (NIR)
A	11,86	1	9,36	2
E	11,15	2	9,15	3
H	10,74	3	9,64	1
D	10,27	4	8,41	5
M	8,00	5	8,63	4
J	6,50	6	6,57	7
C	5,65	7	7,00	6
K	5,28	8	5,59	8
I	5,23	9	5,51	9
Q	5,06	10	4,66	10
G	4,54	11	4,03	13
L	4,52	12	4,61	11
B	3,95	13	4,14	12
N	3,73	14	3,91	14
O	3,45	15	3,52	15
F	3,37	16	2,91	16
P	2,89	17	2,58	17

Tabela 27. Valores obtidos para o *CPD* e *NIR* na *MA*.

Investigador	<i>CPD</i>	Ordenação (<i>CPD</i>)	<i>NIR</i>	Ordenação (<i>NIR</i>)
A	4,79	5	13,85	3
B	2,73	8	8,81	6
C	2,52	9	5,67	9
D	10,17	1	13,53	4
E	1,53	16	5,50	11
F	2,39	10	6,71	8
G	5,36	3	10,17	5
H	2,87	7	7,96	7
I	7,32	2	14,86	1
J	5,21	4	13,91	2
K	3,06	6	5,53	10
L	1,78	13	3,92	15
M	2,38	11	5,09	12
N	1,73	14	3,73	16
O	2,12	12	4,57	13
P	0,79	17	2,07	17
Q	1,63	15	4,22	14

O efeito da normalização torna-se ainda mais evidente quando a ordenação é efetuada considerando todos os investigadores, independentemente da categoria. Os resultados podem ser observados na Tabela 28.

A ordenação dos investigadores utilizando o *CPD* mostra que as posições de topo são ocupadas essencialmente pelos investigadores de *Pharmacology & Pharmacy*. A utilização do *NIR* na ordenação sugere um cenário totalmente diferente em que as posições de topo são agora ocupadas pelos investigadores de *Mathematics, Applied* e nas posições intermédias não existe predominância de investigadores de nenhuma das categorias.

Com o objetivo de melhor entender as diferenças observadas o coeficiente de *Spearman* foi determinado e testadas as seguintes hipóteses:

H_0 : O *CPD* e o *NIR* não estão associados.

H_1 : O *CPD* e o *NIR* estão associados (teste bilateral).

O coeficiente de *Spearman* calculado tem o valor de 0,644 e é significativo ($p < 0,001$). No entanto dado que o valor do coeficiente não evidencia uma correlação forte as seguintes hipóteses foram ainda testadas:

H_0 : O coeficiente de *Spearman* é igual a 0,8.

H_1 : O coeficiente de *Spearman* é inferior a 0,8 (teste unilateral).

Os resultados mostraram que o coeficiente de *Spearman* é inferior a 0,8 e portanto a associação entre os resultados obtidos utilizando o *NIR* e o *CPD* não é forte ($p = 0,04 < 0,05$). O

valor de 0,8 foi estabelecido por se considerar que coeficientes de *Spearman* iguais ou superiores a este valor traduzem uma correlação forte entre as variáveis.

Os testes realizados permitem concluir que o *CPD* e *NIR* fornecem informação diferente e portanto a informação fornecida por exemplo pelo *CPD* não pode ser utilizada para inferir sobre a informação que será fornecida pelo *NIR*. Nesta situação não é possível mencionar que quanto maior o valor de *CPD* maior o valor esperado para o *NIR*.

Tabela 28. Ordenação dos investigadores utilizando o *CPD* e o *NIR*.

Investigador	Categoria	CPD	Ordenação (CPD)	NIR	Ordenação (NIR)
A	PP	11,86	1	9,36	7
E	PP	11,15	2	9,15	8
H	PP	10,74	3	9,64	6
D	PP	10,27	4	8,41	11
D	MA	10,17	5	13,53	4
M	PP	8,00	6	8,63	10
I	MA	7,32	7	14,86	1
J	PP	6,50	8	6,57	15
C	PP	5,65	9	7,00	13
G	MA	5,36	10	10,17	5
K	PP	5,28	11	5,59	17
I	PP	5,23	12	5,51	19
J	MA	5,21	13	13,91	2
Q	PP	5,06	14	4,66	22
A	MA	4,79	15	13,85	3
G	PP	4,54	16	4,03	27
L	PP	4,52	17	4,61	23
B	PP	3,95	18	4,14	26
N	PP	3,73	19	3,91	29
O	PP	3,45	20	3,52	31
F	PP	3,37	21	2,91	32
K	MA	3,06	22	5,53	18
P	PP	2,89	23	2,58	33
H	MA	2,87	24	7,96	12
B	MA	2,73	25	8,81	9
C	MA	2,52	26	5,67	16
F	MA	2,39	27	6,71	14
M	MA	2,38	28	5,09	21
O	MA	2,12	29	4,57	24
L	MA	1,78	30	3,92	28
N	MA	1,73	31	3,73	30
Q	MA	1,63	32	4,22	25
E	MA	1,53	33	5,50	20
P	MA	0,79	34	2,07	34

Poder de previsão do NIR

O poder de previsão do *NIR* foi analisado para dois períodos conjuntamente com o *CPD* e com o *TD*. Na Tabela 29 o *TD*, o *CPD* e o *NIR* estão representados para dois períodos, 1996-2003 e 2000-2007. Os indicadores referidos foram determinados para um conjunto de investigadores da categoria *Pharmacology & Pharmacy*.

Tabela 29. Valores para o *TD*, *CPD* e o *NIR* em 1996-2003 e 2000-2007 na *PP*.

Investigador	1996-2003			2000-2007		
	<i>NIR</i>	<i>CPD</i>	<i>TD</i>	<i>NIR</i>	<i>CPD</i>	<i>TD</i>
H	8,47	11,46	142	9,64	10,74	94
E	7,64	8,43	89	9,15	11,15	85
A	7,50	9,47	173	9,36	11,86	294
I	6,59	6,17	112	5,51	5,23	122
D	6,24	6,79	135	8,41	10,27	134
G	5,13	4,72	57	4,03	4,54	146
M	4,79	4,72	54	8,63	8,00	82
Q	4,31	4,21	122	4,66	5,06	108
F	3,25	3,66	180	2,91	3,37	81
N	2,52	2,23	44	3,91	3,73	91

A amostra de investigadores considerada é a mesma que foi apresentada para o h_{nf} e portanto as mesmas precauções devem ser aqui consideradas relativamente às conclusões retiradas.

Para o período 2000-2007 utilizando o *NIR* verifica-se que a percentagem de pares em que os dois indivíduos mantêm a mesma posição relativamente ao período 1996-2003 é igual á obtida com o indicador *CPD*, 82%. O indivíduo F não influencia os resultados pois ambos os indicadores são independentes da dimensão. O *TD* continua a ser o indicador com menor poder de previsão, 56%. Em relação ao poder de previsão o *NIR* mostrou ser equivalente ao *CPD* e portanto o facto de este ser normalizado torna-o mais adequado para a comparação de indivíduos de diferentes áreas.

Comparação do NIR com o CPP/FCSm e o MNCS

O *crown indicator* (*CPP/FCSm*) e o *mean normalized citation score* (*MNCS*) (Moed, Debruin e van Leeuwen 1995; Waltman, van Eck, van Leeuwen *et al.* 2011a) são indicadores que medem o desempenho científico baseando-se no número de citações. Outros indicadores semelhantes existem, mas apenas estes são aqui mencionados por serem os mais conhecidos na comunidade científica. Assim como no *NIR* também nestes indicadores as citações são normalizadas. O *CPP/FCSm* e o *MNCS* indicam para um conjunto de documentos se o impacto obtido num dado período é superior ou inferior à média mundial determinada para os documentos publicados no mesmo período, do mesmo tipo e em periódicos classificados nas mesmas categorias. Estes indicadores apresentam algumas diferenças relativamente ao *NIR*. O processo de normalização do *NIR* e a informação final obtida são diferentes. Uma comparação do *NIR* com o *CPP/FCSm* e *MNCS* é providenciada.

No *CPP/FCSm* a normalização não é efetuada documento a documento, isto é o aglomerado de citações contabilizadas num dado período são normalizadas como se mostra no exemplo seguinte.

Exemplo

Tabela 30. Informação utilizada no processo de normalização do *CPP/FCSm*.

Documento	Tipo de documento	TC	Categoria	CPD de cada categoria
1	Artigo revisão	13	S1	10,34
2	Artigo	7	S1	5,67
3	Artigo	10	S2	8,79
4	<i>Proceedings paper</i>	2	S3	1,97

Considere-se que os dados apresentados na Tabela 30 pertencem aos documentos publicados por uma dada unidade (apenas a título de exemplo). O *CPP/FCSm* da unidade é calculado da seguinte forma:

$$\frac{CPP}{FCS_m} = \frac{\frac{13 + 7 + 10 + 2}{4}}{\frac{1 * 10,34 + 1 * 5,67 + 1 * 8,79 + 1 * 1,97}{4}} = 1,20 \quad (49)$$

No processo de normalização cada documento deve ser tratado de modo independente pois é de facto uma “peça” de trabalho independente. Este fator foi determinante para preterir o indicador *CPP/FCSm*.

No cálculo do *MNCS* a normalização tem lugar documento a documento. A fórmula de cálculo é a seguinte:

$$MNCS = \frac{1}{n} \times \sum_{i=1}^n \frac{c_i}{e_i} \quad (50)$$

onde n representa o número total de documentos publicados num dado período, c_i é o número de citações atraídas pelo documento i e e_i é a média de citações de todos os documentos do mesmo tipo, publicado no mesmo ano e em periódicos que pertencem à mesma categoria do periódico onde foi publicado o documento i .

Na determinação do *MNCS* se um documento foi publicado num periódico classificado em duas ou mais categorias a média harmónica é aplicada na normalização (Waltman, van Eck, van Leeuwen *et al.* 2011a). No *NIR* a média aritmética é utilizada. Na média harmónica as categorias caracterizadas por uma baixa média de citações por documento contribuem mais para estabelecer a razão entre a média de citações obtidas e a média de citações esperadas. Na aritmética ocorre o comportamento inverso. No entanto considera-se que não existe nenhuma razão plausível que justifique qual das duas médias é a mais indicada pois não é conhecida a verdadeira natureza do documento (domínio científico a que pertence).

O *NIR* é um indicador que fornece informação acerca da média de citações por documento, após a normalização das citações como descrito na secção da metodologia do presente capítulo. A informação que se obtém do *CPP/FCSm* é apenas se o impacto médio de um conjunto de documentos é superior ou inferior à média mundial, mas o seu significado relativamente ao número médio de citações por documento é desconhecido. O *MNCS* também indica se a média de citações de um conjunto de indicadores é superior ou não à média mundial. No exemplo apresentado em seguida, para um caso fictício, é possível ver as diferenças entre a utilização do *MNCS* e do *NIR*.

Exemplo

Tabela 31. Dados utilizados no cálculo do *MNCS* e *NIR*.

Grupo	Tipo de documento	Categoria	TC	$I_{xyi}=e_i$	I_{xy}
G2	Artigo	S1	30	12,01	10,60
	Artigo de revisão	S2	25	21,10	29,19
	<i>Proceedings paper</i>	S3	6	1,20	2,78

O *MNCS* determinado apenas para o artigo de revisão consiste na comparação das citações obtidas por este com as citações esperadas para os artigos de revisão publicados no mesmo ano e em periódicos que pertencem à mesma categoria. No *NIR* o processo de normalização aplicado é diferente. Neste indicador quanto maior o tráfego de citações dentro de uma categoria e para um dado tipo de documento menor é o peso atribuído às citações recebidas por um dado documento. Considere-se os documentos classificados como artigo de revisão e artigo na Tabela 31. O *MNCS* indica que o impacto do artigo (medido pelo número de citações) é superior em cerca de 150% à média dos artigos publicados em periódicos classificados na mesma categoria, enquanto que o artigo de revisão possui um impacto que é apenas superior em 20%. No entanto o comportamento é o inverso quando o *NIR* é calculado devido à valorização do artigo de revisão. Comparando a média de citações por documento para os artigos de S1 com média obtida para todos os artigos publicados no mesmo ano independentemente da categoria é possível verificar que esta categoria é caracterizada por um maior tráfego de citações. Para o caso da categoria S2 verifica-se o oposto pois a média de citações para os artigos de revisão de S2 é inferior à obtida quando se considera a média de citações para todos os documentos deste tipo e publicados no mesmo ano independentemente da categoria a que pertencem. Nesta situação as citações recebidas por este artigo de revisão serão mais valorizadas do que as citações recebidas pelo artigo de S1.

O *MNCS* é um indicador sensível a categorias caracterizadas por uma média de citações muito baixa. Note-se o caso do *proceedings paper* que pertence a um categoria com uma baixa média de citações por documento e que possui um impacto que é cinco vezes superior. No cálculo do *MNCS* para o conjunto dos três documentos este documento contribui consideravelmente para o valor final enquanto que no *NIR* isto não se verifica. O resultado evidencia que o *NIR* é menos sensível a este tipo de cenários.

No *NIR* as publicações mais recentes têm o mesmo peso que as publicações mais antigas. No *MNCS* também se verifica o mesmo. Como os autores indicam isto poderá introduzir algum ruído no indicador (Waltman, van Eck, van Leeuwen *et al.* 2011a). Perante isto os

autores sugerem não considerar no cálculo do indicador o ano mais recente. No *NIR* os resultados obtidos poderão ser testados utilizando um procedimento semelhante.

Considerando os resultados anteriores pode considerar-se que o *NIR* possui os seguintes pontos fortes:

- o *NIR* fornece informação sobre a média de citações por documento depois de normalizadas;
- o processo de normalização aplicado permite a comparação de investigadores estabelecidos em diferentes áreas científicas;
- trata-se de um indicador que apresenta uma sensibilidade limitada às categorias caracterizadas por uma baixa média de citações por documento;
- permite prever o comportamento do investigador no futuro com base na informação disponível no presente. Para um pequeno conjunto de investigadores foi possível prever corretamente a posição relativa dos indivíduos de 82% dos pares disponíveis.

Apesar do conjunto de propriedades que tornam a sua aplicação apropriada quando o objetivo é a comparação de investigadores de diferentes áreas, o *NIR* também apresenta algumas limitações:

- o processo de normalização aplicado torna o indicador sensível a um conjunto de documentos publicados em periódicos classificados simultaneamente em categorias com culturas de citação muito diferentes, devido à utilização da média aritmética;
- após o tratamento individual de cada documento o *NIR* é calculado utilizando a média aritmética. No caso de um pequeno número de documentos muito citados a média obtida poderá ser influenciada consideravelmente. Se no total de documentos existe uma elevada percentagem de documentos com zero citações a média poderá ser igualmente influenciada, mas no sentido da diminuição do seu valor;
- o tipo de normalização adotado, *cited-side*, é dependente do esquema de classificação adotado para os periódicos, neste caso as categorias definidas pelo *WoS*.

3.3.4 Número de autores por documento (*Number of Authors per Document*, NA_m)

O NA_m é um indicador normalizado que permite caracterizar a colaboração científica de um investigador comparando a média de autores por documento, do investigador, com a média de autores por documento que caracteriza as categorias onde tem publicado o investigador. No processo de normalização utilizado foram considerados o tipo de documento, ano de publicação e categoria. A normalização foi efetuada documento a documento. O NA_m representa o número de autores por documento (valor mediano).

As propriedades apresentadas pelo NA_m são discutidas em seguida.

Tipologia dos documentos

As práticas de colaboração diferem entre áreas científicas justificando assim o desenvolvimento de indicadores normalizados que consideram a área científica. Na Tabela 32 é possível verificar algumas diferenças para as categorias do WoS. Apenas três categorias são apresentadas como exemplo.

Tabela 32. Número médio de autores por documento (*APD*) para diferentes tipos de documentos e categorias.

Categoria	Tipo de documentos	APD
<i>Astronomy & Astrophysics</i>	Artigo	7,88
	<i>Proceedings paper</i>	2,79
<i>Mathematics, Applied</i>	Artigo	2,39
	<i>Proceedings paper</i>	2,03
<i>Oncology</i>	Artigo	8,40
	Artigo de revisão	4,19

Os resultados apresentados na Tabela 32 permitem concluir que na categoria *Mathematics, Applied* o número de autores por documento é bastante inferior ao apresentado nas categorias *Astronomy & Astrophysics* e *Oncology*. A *Mathematics, Applied* é uma categoria caracterizada por baixa colaboração entre os investigadores. Os documentos classificados como artigo apresentam um maior número de autores (no caso das categorias aqui utilizadas a título de exemplo). Para a categoria *Mathematics, Applied* não se verifica um comportamento que parece justificar uma normalização por tipo de documentos uma vez que as médias obtidas estão muito próximas para os dois tipos de documentos ao contrário do que se observa para as outras categorias.

O ano de publicação também foi considerado no processo de normalização devido à intensificação nas práticas de colaboração ao longo dos anos (Persson, Glanzel e Danell 2004).

Nas Tabelas 33 e 34 é apresentada uma comparação das ordenações obtidas quando o NA_m e o *APD* são utilizados para descrever a colaboração dos autores de *Pharmacology & Pharmacy* e *Mathematics, Applied*.

Para as duas categorias foi possível observar alterações nas ordenações quando são utilizados o NA_m e o *APD*. Na *Pharmacology & Pharmacy* 47% dos investigadores não sofreram qualquer alteração na ordenação final providenciada pelo NA_m . Para a *Mathematics, Applied* o mesmo comportamento foi verificado para cerca de 7% dos investigadores. A aplicação da normalização parece ser mais importante no caso dos investigadores de *Mathematics, Applied*.

A dispersão dos dados quando o NA_m é utilizado é ligeiramente inferior. Na *Pharmacology & Pharmacy* o coeficiente de variação é de 14% quando o NA_m é usado e 18% quando é considerado o *APD*. Na *Mathematics, Applied* o coeficiente de variação passa de 22% (indicador *APD*) para 20% (indicador NA_m).

Tabela 33. Valores obtidos para o APD e NA_m na PP .

Investigadores	APD	Ordenação (APD)	NA_m	Ordenação (NA_m)
A	7,42	1	6,22	6
M	7,18	2	6,35	4
J	7,14	3	6,36	3
Q	7,04	4	6,58	1
H	6,70	5	6,11	7
D	6,64	6	5,82	8
O	6,56	7	6,40	2
L	6,35	8	6,24	5
E	6,31	9	5,75	9
N	5,68	10	5,47	10
I	5,57	11	5,28	12
G	5,55	12	5,44	11
P	5,31	13	5,23	13
K	5,19	14	4,95	14
C	4,77	15	4,64	15
F	4,11	16	4,30	16
B	3,94	17	4,15	17

Tabela 34. Valores obtidos para o APD e NA_m na MA .

Investigadores	APD	Ordenação (APD)	NA_m	Ordenação (NA_m)
I	3,27	1	6,23	4
J	3,11	2	6,70	1
G	3,05	3	5,68	7
Q	3,00	4	6,52	3
O	3,00	4	6,60	2
C	2,59	5	5,97	6
M	2,56	6	4,89	10
F	2,50	7	6,02	5
K	2,50	7	4,62	12
H	2,40	8	4,96	8
L	2,22	9	4,71	11
D	2,17	10	4,42	13
E	1,97	11	4,94	9
N	1,93	12	4,21	14
B	1,67	13	4,05	15
A	1,64	14	3,76	16
P	1,64	15	3,72	17

O efeito da normalização torna-se ainda mais evidente quando a ordenação é efetuada considerando todos os investigadores, independentemente da categoria. Os resultados podem ser observados na Tabela 35.

Tabela 35. Ordenação dos investigadores utilizando o *APD* e o NA_m .

Investigadores	SC	<i>APD</i>	Ordenação (<i>APD</i>)	NA_m	Ordenação (NA_m)
J	MA	3,11	19	6,70	1
O	MA	3,00	21	6,60	2
Q	PP	7,04	4	6,58	3
Q	MA	3,00	21	6,52	4
O	PP	6,56	7	6,40	5
J	PP	7,14	3	6,36	6
M	PP	7,18	2	6,35	7
L	PP	6,35	8	6,24	8
I	MA	3,27	18	6,23	9
A	PP	7,42	1	6,22	10
H	PP	6,70	5	6,11	11
F	MA	2,50	24	6,02	12
C	MA	2,59	22	5,97	13
D	PP	6,64	6	5,82	14
E	PP	6,31	9	5,75	15
G	MA	3,05	20	5,68	16
N	PP	5,68	10	5,47	17
G	PP	5,55	12	5,44	18
I	PP	5,57	11	5,28	19
P	PP	5,31	13	5,23	20
H	MA	2,40	25	4,96	21
K	PP	5,19	14	4,95	22
E	MA	1,97	28	4,94	23
M	MA	2,56	23	4,89	24
L	MA	2,22	26	4,71	25
C	PP	4,77	15	4,64	26
K	MA	2,50	24	4,62	27
D	MA	2,17	27	4,42	28
F	PP	4,11	16	4,30	29
N	MA	1,93	29	4,21	30
B	PP	3,94	17	4,15	31
B	MA	1,67	30	4,05	32
A	MA	1,64	31	3,76	33
P	MA	1,64	32	3,72	34

A utilização do indicador *APD* mostra que as posições de topo são ocupadas pelos investigadores de *Pharmacology & Pharmacy*. O cenário altera-se quando o NA_m é aplicado. Nas posições de topo não existe predominância de nenhuma das categorias, mas a primeira e segunda posição passam a ser ocupadas pelos investigadores de *Mathematics, Applied*.

Com o objetivo de melhor entender as diferenças observadas o coeficiente de *Spearman* foi determinado e as seguintes hipóteses testadas:

H_0 : O *APD* e o NA_m não estão associados.

H_1 : O *APD* e o NA_m estão associados (teste bilateral).

O valor determinado para o coeficiente de *Spearman* tem o valor de 0,616 e é significativo ($p < 0,001$), logo os coeficientes estão associados. Uma vez que o valor da correlação não evidencia uma correlação forte as seguintes hipóteses foram ainda testadas:

H_0 : O coeficiente de *Spearman* é igual a 0,8

H_1 : O coeficiente de *Spearman* é inferior a 0,8 (teste unilateral)

Os resultados mostraram que o coeficiente de *Spearman* é inferior a 0,8 e portanto a associação entre os resultados obtidos utilizando o *APD* e o NA_m não é forte ($p = 0,02 < 0,05$).

Os resultados alcançados nos testes permitem concluir que a informação fornecida pelo *APD* e NA_m não é a mesma, isto é, o aumento do valor de *APD* não implica um aumento de NA_m e vice-versa.

Poder de previsão do NA_m

O poder de previsão do NA_m foi analisado para dois períodos em conjunto com o *APD* e o *TD*. Na Tabela 36 estão representados os três indicadores para os períodos 1996-2003 e 2000-2007. Os indicadores referidos foram determinados para um conjunto de investigadores da categoria *Pharmacology & Pharmacy*.

Tabela 36. Valores para o *TD*, *APD* e NA_m em 1996-2003 e 2000-2007 na *PP*.

Investigador	1996-2003			2000-2007		
	NA_m	<i>APD</i>	<i>TD</i>	NA_m	<i>APD</i>	<i>TD</i>
A	6,37	7,91	173	6,22	7,42	294
M	6,21	7,04	54	6,35	7,18	82
Q	5,94	6,36	122	6,58	7,04	108
H	5,64	6,09	142	6,11	6,70	94
D	5,22	6,12	135	5,82	6,64	134
N	5,17	5,45	44	5,47	5,68	91
I	5,04	5,20	112	5,28	5,57	122
E	4,77	5,25	89	5,75	6,31	85
G	4,58	4,63	57	5,44	5,55	146
F	4,22	4,12	180	4,30	4,11	81

A amostra de investigadores utilizada para testar o poder de previsão do indicador é a mesma que se usou para o h_{nf} e *NIR* e portanto as recomendações relativas às conclusões retiradas devem ser mantidas.

A comparação da posição dos dois investigadores que constituem os diversos pares disponíveis na amostra no período 1996-2003 e 2000-2007 mostrou que quando o NA_m é

utilizado a percentagem de pares cujos investigadores mantêm as posições em 2000-2007 relativamente a 1996-2003 é de 87%. Se o *APD* é utilizado é possível prever corretamente a posição dos investigadores em 91% do total dos pares para o período 2000-2007 utilizando a informação que se dispõem em 1996-2003. O *TD* continua a ser o indicador com menor poder de previsão, permitindo prever, corretamente em 2000-2007, apenas a posição dos investigadores de 56% dos pares.

Os valores obtidos para a percentagem de pares previstos corretamente utilizando o *APD* e o *NA_m* não estão muito distantes. A normalização introduzida no *NA_m* será certamente um fator determinante na escolha deste indicador para descrever a colaboração de investigadores que trabalham em diferentes áreas científicas.

Considerando os resultados anteriores pode considerar-se que o *NA_m* possui os seguintes pontos fortes:

- o *NA_m* fornece informação sobre o número de autores por documento (valor mediano) após a normalização;
- o processo de normalização aplicado permite a comparação de investigadores estabelecidos em diferentes áreas científicas;
- permite prever o comportamento de investigadores no futuro com base na informação disponível no presente. Para um pequeno conjunto de investigadores foi possível prever corretamente a posição relativa dos investigadores de 87% dos pares disponíveis;
- após o tratamento individual de cada documento o *NA_m* representa o valor mediano da distribuição que representa o conjunto de documentos de cada investigador. Documentos com um número de autores atípico não representam um problema pois a mediana não é afetada pela presença dos mesmos.

Apesar do conjunto de propriedades que tornam a sua aplicação apropriada quando o objetivo é a comparação de investigadores de diferentes áreas, o *NA_m* também apresenta algumas limitações:

- o processo de normalização aplicado torna o indicador sensível a um conjunto de documentos publicados em periódicos classificados simultaneamente em categorias com culturas de colaboração muito diferentes, devido à utilização da média aritmética;
- o tipo de normalização adotado é dependente do esquema de classificação adotado para os periódicos, neste caso as categorias definidas pelo *WoS*

3.4 Conclusões

Neste capítulo foram discutidos os indicadores desenvolvidos e foi efetuada uma comparação com outros já existentes com o objetivo de evidenciar as vantagens de cada indicador e as suas limitações.

Nesta secção apenas se apresentam as conclusões gerais e sugestões para trabalho futuro.

Ao longo do trabalho procurou desenvolver-se indicadores que possam ser aplicados na comparação do desempenho de investigadores de diferentes áreas científicas. Todos os indicadores apresentados envolvem a normalização de citações com exceção do NA_m . O processo de normalização utilizado foi o *cited-side*. No entanto, estudos recentes sugerem que a normalização baseada no *citing-side* poderá ter algumas vantagens. A utilização desta metodologia no cálculo de indicadores ao nível individual deverá ser explorada em estudos futuros.

Para todos os indicadores foi identificado um conjunto de propriedades que os torna atrativos na descrição do desempenho. No entanto, estes também possuem algumas limitações que não podem ser esquecidas no momento da interpretação dos resultados obtidos.

Os indicadores estudados exigem que uma grande quantidade de dados seja recolhida para que o processo de normalização tenha lugar. O processo de recolha da informação é bastante moroso quando se recorre apenas aos instrumentos disponibilizados *on-line* pela base de dados utilizada.

As auto-citações não foram eliminadas no cálculo dos indicadores. Este é um assunto muito discutido entre os bibliómetros. Contudo, não existe um consenso sobre o verdadeiro papel das auto-citações no impacto alcançado pelos documentos. Em estudos futuros deverá ser explorado o efeito das auto-citações nos resultados fornecidos pelos indicadores aqui apresentados. Espera-se que a influência seja mínima pois consideraram-se longas janelas de citação (comprimento máximo de 10 anos) e, tal como o estudo efetuado por Aksnes 2003 mostra, o efeito das auto-citações no impacto obtido por um conjunto de documentos diminui após o terceiro ano de publicação.

Capítulo 4

Definição do Modelo

De entre o conjunto de parâmetros analisados nos concursos o desempenho científico foi o utilizado na construção de um modelo que se pretende ser capaz de servir de apoio à decisão neste tipo de avaliações. Para descrever o desempenho científico foram utilizados vários indicadores bibliométricos que permitem caracterizar as suas diversas dimensões.

A decisão final do júri em cada concurso é a ordenação dos candidatos que resulta da avaliação de um conjunto de parâmetros. Ao melhor candidato é atribuída a posição 1, ao segundo melhor candidato a posição 2 e assim sucessivamente até não existirem mais candidatos. Esta ordenação foi utilizada como variável dependente, isto é, como variável que se pretende reproduzir a partir de indicadores bibliométricos.

Os documentos publicados por um candidato são o ponto de partida na avaliação do desempenho científico. A avaliação poderá revelar-se uma tarefa muito morosa e complexa, dependendo da área científica e do percurso científico do candidato. Numa situação deste género a informação fornecida pelos indicadores bibliométricos ajudará o avaliador na formulação de um juízo acerca do candidato.

As decisões de cada painel não são apenas influenciadas pelo desempenho científico. Outros parâmetros são também avaliados, tais como o desempenho pedagógico, as atividades de gestão e de transferência de tecnologia. No entanto, não foi possível obter informação relacionada com estes parâmetros. Dado que para a maioria dos concursos aqui considerados o desempenho científico é a vertente que possui maior peso espera encontrar-se alguma relação entre os indicadores bibliométricos e as decisões finais dos concursos.

O processo de seleção nestes concursos ou noutros que envolvam a avaliação do desempenho científico, não pode ser efetuado utilizando unicamente um modelo baseado em indicadores bibliométricos. Se a definição de um modelo baseado em indicadores bibliométricos for viável ele poderá ser usado como um instrumento auxiliar na avaliação do desempenho científico e de apoio à decisão.

No presente capítulo descreve-se o modelo desenvolvido.

4.1 Amostra

No exercício de construção dos modelos foi utilizada uma amostra constituída por 174 investigadores que apresentam as seguintes características:

- possuem como afiliação uma instituição de ensino superior portuguesa;
- todos têm no mínimo 10 anos de atividade científica após a finalização do doutoramento;
- foram candidatos a um lugar de Professor Associado ou Catedrático numa das universidades portuguesas;
- o número mínimo de documentos publicados e indexados no WoS nos 10 anos que antecedem a data do concurso é de 10;
- foram candidatos a concursos na área da química, física, biologia, matemática, geologia e engenharia.

A produção científica de cada candidato nos 10 anos que antecederiam a data do lançamento do concurso foi a única considerada neste trabalho por duas razões:

- considera-se que os membros do painel prestarão mais atenção e estarão mais interessados nas atividades de investigação de cada candidato e dos seus resultados no período mais recente;
- para todos os candidatos da amostra decorreram no mínimo 10 anos após o ano de doutoramento.

4.2 Metodologia

4.2.1 Pesquisa da produção científica no *WoS* para investigadores

A recolha da informação relativa à produção científica dos investigadores foi efetuada no *WoS*. A metodologia aplicada foi a descrita no Capítulo 3 utilizada na amostra de investigadores holandeses.

4.2.2 Análise estatística

O modelo de regressão de escolha discreta foi utilizado na modelação das decisões tomadas pelos painéis de avaliação, já que é o mais adequado para trabalhar dados que apresentam as características da variável dependente em análise.

Estes modelos serão mais eficientes se as variáveis independentes não se encontrarem correlacionadas. Nestas condições, foi avaliada a multicolineariedade entre os indicadores bibliométricos selecionados (variáveis independentes) utilizando o “fator de inflação da variância” (*VIF*) e o “índice de condição” (*CI*) (Maroco 2007).

Problemas de multicolineariedade foram de facto detetados como se mostrará posteriormente. Para eliminar (ou reduzir significativamente) a multicolineariedade foram estudadas três situações alternativas:

- Estratégia 1 - Utilizar a análise fatorial;
- Estratégia 2 - Eliminar o indicador h_{mf} do conjunto de indicadores por estar fortemente correlacionado com o *NDF*;
- Estratégia 3 - Eliminar o indicador *NDF* do conjunto de indicadores.

A adoção das três estratégias de acordo com o procedimento que a seguir se descreve, levou à definição de três modelos distintos.

4.2.2.1 Análise fatorial

Tal como já foi anteriormente referido, a análise fatorial foi utilizada para eliminar a multicolineariedade entre as variáveis independentes. A análise fatorial é uma técnica multivariada com a qual se procura reduzir o universo de variáveis disponíveis, tentando encontrar-se uma estrutura subjacente mais simples. Deste modo, o objetivo da análise fatorial passa por condensar a informação contida no conjunto de variáveis originais (observáveis) num número reduzido de dimensões (fatores) com a menor perda de informação possível. Esta operação é efetuada tendo por base as correlações entre as variáveis originais.

Na análise fatorial cada variável x é expressa como uma combinação linear dos fatores chamados comuns (F_i) e de um fator que é único à variável (U_j).

$$x_1 = b_{11}F_1 \dots b_{1k}F_k + U_1 \quad (51)$$

...

$$x_p = b_{p1}F_1 \dots b_{pk}F_k + U_p \quad (52)$$

O fator único associado a cada uma das variáveis representa a variabilidade que não é explicada pelos fatores comuns.

Os coeficientes b_{ji} , $j=1, \dots, p$ e $i=1, \dots, k$ são os pesos fatoriais que representam as correlações entre as variáveis observáveis e os fatores comuns. Para cada variável, o somatório dos quadrados dos pesos fatoriais traduz a variância total explicada pelos fatores comuns. Esta variância é denominada por “comunalidade” e é expressa pela seguinte equação:

$$h_p^2 = (b_{p1})^2 + \dots + (b_{pk})^2 \quad (53)$$

Vários critérios devem ser respeitados antes de se proceder à utilização desta técnica. Tais critérios e as metodologias utilizadas na análise que foi efetuada descrevem-se em seguida.

Tamanho da amostra

O tamanho da amostra deve ser suficiente para permitir que a aplicação da análise fatorial seja válida. O número mínimo de observações na amostra depende do número de variáveis consideradas. Não existe consenso quanto a este valor. Alguns autores consideram que aquele valor deve ser no mínimo cinco vezes o número de variáveis disponíveis, outros defendem que deve ser dez vezes e existem mesmo aqueles que consideram que o número mínimo deve ser vinte (Hair 2010). No caso em estudo o tamanho da amostra respeita os requisitos mencionados na literatura.

Dados omissos

Pretende-se que exista uma baixa percentagem de dados omissos na amostra (inferior a 20% do total de dados). No caso em apreço, no conjunto de dados disponíveis não existem valores omissos.

Valores atípicos

Também não é desejável a existência de observações que se distanciam de forma significativa das restantes. A identificação de observações atípicas foi efetuada neste estudo usando diagramas de extremos e quartis e a distância de *Mahalanobis* (Barnett e Lewis 1994).

Normalidade

Na análise fatorial os pesos fatoriais e as variâncias são determinados de modo a que ocorra a menor perda possível de informação. Nesse processo podem utilizar-se vários métodos de estimação. Os mais praticados são o método das componentes principais, o método da factorização do eixo principal e o método da máxima verosimilhança. Apenas o último exige que os dados sigam uma distribuição Normal (Maroco 2007). Tal como se verá posteriormente, o método de estimação utilizado neste trabalho foi o das componentes principais. Apesar de este método não exigir a normalidade dos dados, cada uma das variáveis foi analisada individualmente em relação à assimetria da sua distribuição, já que distribuições muito assimétricas podem distorcer os resultados. Nos casos em que a transformação da variável permitiu obter uma distribuição Normal foi utilizada a variável transformada. Para o estudo da normalidade recorreu-se ao teste de *Shapiro-Wilk* (*Shapiro e Wilk 1965*).

Matriz de correlações

A realização da análise fatorial apenas faz sentido se existir um número suficiente de variáveis correlacionadas entre si. Assim, determinou-se a matriz de correlações e verificou-se se a mesma era adequada para a aplicação da análise fatorial. A adequação da matriz de correlações foi efetuada recorrendo:

- ao teste de esfericidade de *Bartlett*. Este teste determina se a matriz de correlações é a matriz identidade. Em caso afirmativo a análise fatorial não poderá ser aplicada;
- à medida de adequação da amostra de *Kaiser-Meyer-Olkin* (*KMO*). O *KMO* varia entre 0 e 1 e valores próximos de 1 indicam que cada variável pode ser prevista utilizando as restantes variáveis, com um baixo erro associado;

$$KMO = \frac{\sum_{j=1}^{p-1} \sum_{m=j+1}^p r_{x_j x_m}^2}{\sum_{j=1}^{p-1} \sum_{m=j+1}^p r_{x_j x_m}^2 + \sum_{j=1}^{p-1} \sum_{m=j+1}^p r_{x_j x_m | x_n}^2} \quad (54)$$

Na expressão (54) $r_{x_j x_m / x_n}^2$ representa o quadrado da correlação parcial entre as variáveis x_j e x_m quando o efeito da variável x_n é eliminado. A expressão permite calcular o *KMO* para toda a amostra. O *KMO* pode também ser calculado para uma variável específica após uma pequena alteração da expressão:

$$KMO_j = \frac{\sum_{m \neq j=1}^p r_{x_j x_m}^2}{\sum_{m \neq j=1}^p r_{x_j x_m}^2 + \sum_{m \neq j=1}^p r_{x_j x_m / x_n}^2} \quad (55)$$

Quando o *KMO* se refere a uma variável, a sigla que aparece associada a este parâmetro nos *softwares* estatísticos é *MSA* (*Measure of Sampling Adequacy*) e será esta a que será utilizada ao longo deste trabalho.

O *KMO* foi determinado para toda a mostra e para cada uma das variáveis.

- coeficiente de correlação parcial. A aplicação da análise fatorial é viável se os coeficientes de correlação parciais forem pequenos. O coeficiente de correlação parcial é determinado usando a seguinte expressão:

$$r_{x_j x_m / x_n} = \frac{(r_{x_j x_m} - r_{x_j x_n} r_{x_m x_n})}{\sqrt{(1 - r_{x_j x_m}^2)(1 - r_{x_m x_n}^2)}} \quad (56)$$

Métodos de extração de fatores

Após a análise da matriz de correlações procedeu-se à extração dos fatores. Os três métodos mais utilizados são: componentes principais, o método da factorização do eixo principal e o método da máxima verosimilhança. O método da verosimilhança foi preterido já que nem todas as variáveis analisadas eram normais (mesmo quando foi efetuada uma transformação) e o método exige que as variáveis possuam uma distribuição Normal. O método da factorização do eixo principal não foi escolhido por ser o mais afetado pela indeterminação dos fatores, isto é, várias estruturas fatoriais poderem ser obtidas a partir dos mesmos dados (Hair 2010).

O método das componentes principais foi assim o selecionado. Este método utiliza um processo iterativo na determinação dos pesos fatoriais e das comunalidades. Para determinar os pesos fatoriais é necessário dispor do valor das comunalidades. Sucede que para determinar o valor das comunalidades é necessário conhecer os pesos fatoriais. Isto gera um problema de referências circulares que obriga à utilização de um processo iterativo. O primeiro passo deste processo é fixar o valor da comunalidade $h_j^2 (j = 1, \dots, p)$ que representa o valor máximo da correlação da variável com as restantes variáveis. O valor inicial para h_j^2 é 1 assumindo-se assim que toda a variância da variável é explicada pelas restantes variáveis e que, portanto, o número de fatores comuns é igual ao número total de variáveis. Partindo deste ponto e de acordo com a regra do valor próprio superior a 1 (critério de *Kaiser*), um determinado número de fatores irá ser retido. O primeiro será aquele que explica uma maior percentagem da variância das variáveis consideradas; o segundo possui associado uma menor

percentagem em relação ao primeiro mas superior ao terceiro fator e assim sucessivamente até não existirem fatores com valores próprios superiores a 1.

No final os pesos fatoriais obtidos constituem uma medida de associação dos fatores com cada variável. A matriz de pesos fatoriais obtidos é representada por:

$$L = \begin{bmatrix} b_{11} & \dots & b_{1k} \\ \vdots & \dots & \vdots \\ b_{p1} & \dots & b_{pk} \end{bmatrix}$$

Tal como anteriormente mencionado, os pesos fatoriais permitem determinar a variância de cada variável que é explicada pelos fatores comuns.

Método de rotação de fatores

O objetivo da rotação dos fatores é melhorar a interpretação da estrutura fatorial obtida. Antes da rotação o conjunto de pesos fatoriais pode não permitir uma fácil interpretação da estrutura fatorial. Com a rotação procura-se que o peso fatorial de uma variável seja o maior possível num dos fatores e o menor possível nos restantes fatores.

Existem dois métodos de rotação de fatores; o método ortogonal e o método oblíquo. No método ortogonal os eixos são mantidos a 90°, enquanto que na rotação oblíqua após rotação os eixos podem ser mantidos a qualquer distância. Aqui foi adotada a rotação ortogonal porque se pretendia obter no final um conjunto de fatores que não estivessem correlacionados entre si. É importante relembrar que a análise fatorial foi aplicada com o objetivo de eliminar a multicolineariedade. A aplicação de uma rotação oblíqua não garante a independência dos fatores.

Para a rotação ortogonal dos fatores existem também vários métodos. Os mais populares são a rotação *Varimax* e a rotação *Quartimax*. A rotação *Varimax* é sem dúvida a mais utilizada. A rotação *Quartimax* deve ser utilizada quando o analista suspeita que a estrutura fatorial pode ser explicada por um fator mais geral e por outros fatores específicos. No caso em estudo não se conhecia nada acerca da estrutura fatorial e, por isso, foi adotada a rotação *Varimax*.

Na rotação *Varimax* o objetivo é encontrar uma matriz ortogonal à matriz dos pesos fatoriais de modo a que a variância dos quadrados dos pesos fatoriais para cada variável seja máxima (com a condição de que as communalidades devem ser as mesmas antes e após a rotação). Nesta rotação, os pesos fatoriais representam coeficientes de correlação de *Pearson* entre os fatores e as variáveis.

Número de fatores retidos

Na decisão sobre o número de fatores a reter foi utilizado o critério de *Kaiser*. Isto é, foram retidos todos os fatores com valores próprios superiores a 1. No entanto, outras metodologias de seleção dos fatores foram aplicadas, mas não foram encontrados resultados superiores.

Análise da estrutura fatorial

A estrutura fatorial que resultou foi estudada detalhadamente. Os pesos fatoriais obtidos foram analisados com o objetivo de identificar quais as variáveis que eram determinantes na definição de cada fator. A significância dos pesos fatoriais foi também analisada.

Foi examinada a proporção da variância de cada variável explicada pela estrutura fatorial obtida. Nesta análise foi imposto o requisito mínimo de 50% para os valores das comunalidades.

A proporção da variância total explicada pelos fatores retidos foi também objeto de estudo. Valores muito baixos não são satisfatórios uma vez que o objetivo é identificar um novo conjunto de variáveis, em menor número que o conjunto inicial, com a menor perda possível de informação.

Após a análise destes parâmetros prosseguiu-se com a determinação dos valores dos *factor scores*.

Estimação dos factor scores

Dado que o objetivo da análise fatorial é compreender o comportamento de um dado fenómeno com base num conjunto de fatores subjacentes, fica então concluída no momento em que a estrutura fatorial é definida. Nos casos em que se pretende utilizar os resultados em análises posteriores é necessário proceder à caracterização das novas variáveis que são representadas por cada um dos fatores retidos. Para este objetivo existem dois procedimentos alternativos:

- a) selecionar em cada fator a variável com o maior peso fatorial e considerar esta como representativa do fator. Apenas esta variável será usada nas análises posteriores.
- b) as variáveis originais são substituídas por um menor número de variáveis. As novas variáveis são variáveis compostas.

Relativamente à primeira alternativa, a seleção de uma única variável por fator pode revelar-se uma tarefa difícil, pois podem existir no mesmo fator variáveis com pesos fatoriais elevados semelhantes. Na segunda situação as variáveis compostas podem ser construídas utilizando o somatório de escalas ou *factor scores*. Na elaboração do somatório de escalas apenas as variáveis com elevados pesos fatoriais em cada fator são consideradas. Neste caso a independência entre as variáveis compostas obtidas não é garantida. Nos *factor scores* todas as variáveis do fator são consideradas e a relação da variável com o fator é tida em consideração.

Na determinação dos *factor scores* existem três métodos: o método de *Bartlett*, o método de *Thompson* e o método de *Anderson-Rubin*. O último método foi adotado por ser o único que garante que os *factor scores* não estão correlacionados.

Este método fornece um conjunto de coeficientes que permitem determinar o *score* para cada fator:

$$C = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \dots & \vdots \\ a_{p1} & \dots & a_{pk} \end{bmatrix}$$

$$FS_I = a_{1I}x_I + \dots + a_{pI}x_p \quad (57)$$

Para um fator, o *factor score*, FS_I , resulta da combinação linear das variáveis observáveis de acordo com a contribuição de cada uma para a definição do fator. A principal desvantagem dos *factor scores* é a dificuldade em replicar os resultados entre estudos, pois estes são obtidos a partir da matriz dos pesos fatoriais e a obtenção de matrizes iguais entre estudos é difícil.

Análise da fiabilidade

O conceito de fiabilidade, ou consistência interna, considera que as variáveis que definem um fator devem medir o mesmo objeto e, portanto, devem estar correlacionadas entre si. A fiabilidade foi estudada utilizando o *Cronbach's α* que é determinado pela expressão:

$$Cronbach's \alpha = \frac{p^2 \mu_{Cov}}{\sum_{j=1}^p \sigma_j^2 + \sum_{j=1}^p Cov_j} \quad (58)$$

O cálculo do *Cronbach's α* requer que a matriz de variância-covariância seja determinada considerando todas as variáveis. Os valores da diagonal desta matriz representam a variância de uma variável e os valores fora da diagonal representam a covariância entre duas variáveis. Na expressão 58 o numerador representa o número total de variáveis ao quadrado multiplicado pela covariância média de todas as variáveis. O denominador corresponde à soma da variância e covariância de todas as variáveis.

Valores de *Cronbach's α* no intervalo [0,7-0,8] são satisfatórios (embora valores inferiores são considerados razoáveis em algumas situações, principalmente se existe muita heterogeneidade). No entanto, quando o *Cronbach's α* é utilizado para analisar a consistência interna é importante ter em atenção que este índice depende do número de variáveis ao quadrado. Perante uma análise em que o número de variáveis é elevado o *Cronbach's α* pode ser elevado não porque existe fiabilidade mas sim devido ao elevado número de variáveis presentes (Field 2005).

O *Cronbach's α* foi calculado para cada um dos fatores retidos.

4.2.2.2 Modelos de escolha discreta

Após a aplicação da análise fatorial recorreu-se a um modelo de escolha discreta para atingir o objetivo pretendido.

Os modelos de escolha discreta são muito utilizados em econometria e *marketing*. A primeira aplicação deste tipo de modelos ocorreu em 1962 no estudo da escolha modal entre transporte público e privado (Warner 1962).

O objetivo dos modelos de escolha discreta é o de prever escolhas efetuadas por um indivíduo que perante um conjunto de alternativas finitas tem de optar por uma delas. Tais modelos aplicam-se, por exemplo, quando está em causa prever a compra de um automóvel da marca A, B ou C.

Existem vários modelos de escolha discreta dependendo das características da variável dependente disponível:

- *Probit regression* e *logit regression* - variável dependente binária.
- *Ordered probit regression* e *ordered logit regression* - variável dependente ordinal.
- *Multinomial probit regression* e *multinomial logit regression* - variável dependente nominal
- *Bivariate probit regression* - variável dependente bivariada
- *Rank-ordered probit regression* e *rank-ordered logit regression* – variável dependente em forma ordenada (*ranking*).

Nos concursos considerados neste estudo o painel de avaliação produz a decisão final de uma forma ordenada, isto é, a cada candidato é atribuída uma posição. A ordenação é decrescente ocupando a posição um o melhor candidato. De entre o conjunto de modelos de escolha discreta o *rank-ordered logit regression (ROLR)* é aquele que permite trabalhar com este tipo de variável dependente. Nestas condições, apenas o *logit* e o *ROLR* irão ser descritos.

Características das alternativas de escolha

As alternativas disponíveis para um decisor devem ser mutuamente exclusivas, exaustivas e finitas para que seja possível a aplicação de modelos de escolha discreta.

No estudo em causa as alternativas são mutuamente exclusivas e exaustivas pois o júri não pode atribuir a mesma posição a dois candidatos e a todos os candidatos tem de ser atribuída uma posição na ordenação final.

Construção do modelo

O *logit* e o *rank-ordered logit* baseiam-se na Teoria da Utilidade Estocástica, isto é, a cada alternativa está associada uma utilidade do ponto de vista do decisor e que é desconhecida antes da decisão. Perante um conjunto de decisores considera-se que existe variabilidade de preferências. Nestas condições, o decisor irá efetuar a sua escolha em função da utilidade que associa a cada alternativa: perante duas alternativas escolherá aquela que, do seu ponto de vista, tem a maior utilidade associada.

Se a utilidade for representada por U_{nj} , em que n representa o decisor e $j = 1, 2, \dots, J$ as alternativas, o decisor apenas irá escolher a alternativa i se:

$$U_{ni} > U_{nj} \quad j \neq i \quad (59)$$

A utilidade que o decisor atribui a uma determinada alternativa apenas é conhecida no seu todo pelo mesmo.

O analista que procura explicar as escolhas feitas pelos decisores perante um conjunto de alternativas apenas pode observar alguns dos atributos pertencentes às alternativas, x_{nj} , e aos decisores, s_n . Com base nestas variáveis é possível especificar uma função que relaciona os atributos observados com a utilidade que o decisor atribui a uma dada alternativa. Esta função é denominada como utilidade representativa e é expressa por:

$$V_{nj} = V(x_{nj}, s_{nj}) \quad \forall j \quad (60)$$

Devido à existência de atributos que não podem ser observados pelo analista a utilidade de cada alternativa resulta da adição de:

- um termo determinístico, V_{nj} , que é função dos atributos da utilidade observados pelo analista;
- um termo estocástico que o analista trata como aleatório, ε_{nj} , com uma dimensão que depende do rigor da informação prévia e da variabilidade das preferências dos decisores. Este termo, que não é observável pelo analista, é muitas vezes denominado de erro.

A utilidade é então representada por:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad (61)$$

Na maioria das situações o termo determinístico resulta da combinação linear dos atributos observáveis pelo analista. Este termo é representado por:

$$V_{nj} = \beta x_{nj} \quad (62)$$

O vetor das variáveis observadas é representado por x_{nj} e o vetor β representa o contributo de cada uma das variáveis.

O modelo *logit* é definido considerando um conjunto de hipóteses sobre o termo aleatório da utilidade. As hipóteses consideradas são as seguintes:

- os erros são independentes e possuem todos a mesma distribuição com os mesmos parâmetros (são independentemente e identicamente distribuídos – *idd*);
- os erros possuem uma distribuição de valor extremo, a distribuição de *Gumbel*;
- os erros assumem valores pequenos relativamente ao termo determinístico.

A densidade de probabilidade é dada por:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \quad (63)$$

e a função distribuição é:

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}} \quad (64)$$

Conhecendo a densidade de probabilidade o analista pode inferir sobre a probabilidade associada à escolha feita pelo decisor. A probabilidade de um decisor n escolher a alternativa i é dada por:

$$P_{ni} = P(U_{ni} > U_{nj}, \forall j \neq i) \quad (65)$$

$$P_{ni} = P(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni}, \forall j \neq i) \quad (66)$$

Esta expressão representa a probabilidade de cada termo aleatório ser inferior ao termo determinístico. A distribuição acumulada desta probabilidade é dada por:

$$P_{ni} = \int I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (67)$$

O termo $I(\cdot)$ é uma função indicadora. O valor desta função é 1 quando o argumento é verdadeiro e 0 quando o argumento é falso.

Esta probabilidade é um integral de J dimensões sobre a densidade dos erros. Considerando que apenas a diferença entre utilidades é importante que o número de diferenças entre os termos do erro é $J-1$, é então possível expressar esta probabilidade a $J-1$ dimensões sobre a densidade das diferenças entre os termos erro:

$$P_{ni} = \int I(\varepsilon_{nji} < V_{ni} - V_{nj}, \forall j \neq i) g(\varepsilon_{ni}) d\varepsilon_{ni} \quad (68)$$

A irrelevância da escala da utilidade

A alternativa com maior utilidade irá ser sempre a mesma, independentemente da escala da utilidade. Se, por exemplo, a utilidade $U_{ni} = V_{ni} + \varepsilon_{ni} \forall i \neq j$ for multiplicada por uma constante λ ($U_{ni}^1 = \lambda \times V_{ni} + \lambda \times \varepsilon_{ni}$ e $\lambda > 0$) as escolhas efetuadas pelo decisor mantêm-se. É neste sentido que a normalização das utilidades é necessária. A normalização da escala da utilidade é efetuada normalizando os termos do erro. A escala da utilidade e a variância dos erros estão relacionadas. Quando a utilidade é multiplicada por λ a variância de cada ε_{ni} passa a ser: $\lambda^2 \text{var}(\varepsilon_{ni})$. Deste modo normalizar a variância do termo erro é o mesmo que normalizar a escala da utilidade. As variâncias dos erros no modelo *logit* padrão são normalizadas para $\frac{\pi^2}{6}$, $\text{var}(\varepsilon_{ni}) = 1,64$ (Train 2009, Capítulo 3).

Então o modelo passa a ser definido pela seguinte expressão:

$$U_{ni} = \frac{\beta^*}{\sigma} \times \frac{\pi}{\sqrt{6}} \times x_{ni} + \varepsilon_{ni} \quad (69)$$

Se a variância:

$$\text{var}(\varepsilon_{ni}) = \frac{\pi^2}{6} \quad (70)$$

então:

$$\lambda^2 \text{var}(\varepsilon_{ni}) = \frac{\pi^2}{6} \quad (71)$$

$$\lambda = \frac{\pi}{\sqrt{6} \times \sigma} \quad (72)$$

Na determinação dos parâmetros do modelo apenas $\frac{\beta^*}{\sigma} \times \frac{\pi}{\sqrt{6}}$ pode ser determinada e portanto β representa este valor. O valor de β^* e σ não podem ser determinados independentemente. Aqui o valor de β deve ser interpretado relativamente à variância dos erros que constituem os atributos não observáveis.

Considerando que apenas a diferença entre as utilidades é relevante, que a normalização ocorre automaticamente e que ε_{nj} são independentes, a função distribuição conjunta de todos as $j \neq i$ resulta do produto das funções de distribuição individuais:

$$P_{ni}/\varepsilon_{ni} = \prod_{j \neq i} e^{-e^{(\varepsilon_{ni} + V_{ni} - V_{nj})}} \quad (73)$$

Uma vez que o valor de ε_{ni} não é conhecido a solução passa pelo cálculo do valor da probabilidade usando o integral de P_{ni}/ε_{ni} sobre todos os valores de ε_{ni} e considerando as respetivas densidades:

$$P_{ni} = \int \prod_{j \neq i} e^{-e^{(\varepsilon_{ni} + V_{ni} - V_{nj})}} e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \quad (74)$$

A integração desta expressão permite obter para cada alternativa a probabilidade de ser seleccionada em primeiro lugar:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \quad (75)$$

Substituindo as utilidades resulta que:

$$P_{ni} = \frac{e^{\beta \cdot x_{ni}}}{\sum_j e^{\beta \cdot x_{nj}}} \quad (76)$$

O numerador desta expressão representa a exponencial da utilidade da alternativa i e o denominador a soma de todas as exponenciais das utilidades do conjunto de alternativas disponíveis.

As probabilidades determinadas utilizando o modelo *logit* possuem as seguintes características:

- tal como seria necessário, as probabilidades pertencem ao intervalo]0-1[. A probabilidade P_{ni} aproxima-se de 1 quando a utilidade associada à alternativa i aumenta e as utilidades das restantes alternativas se mantêm constantes. A probabilidade aproxima-se de 0 se a utilidade associada à alternativa i diminuir e as utilidades associadas as restantes alternativas se mantiverem constantes;
- a soma das probabilidades do conjunto de alternativas é igual a 1;
- a relação entre a probabilidade determinada e a componente determinística da utilidade é uma sigmoide. Perante esta relação, um aumento da utilidade terá maior efeito (de aumento) sobre a probabilidade se esta estiver próxima do valor 0,5. Isto pode ser observado na Figura 10.

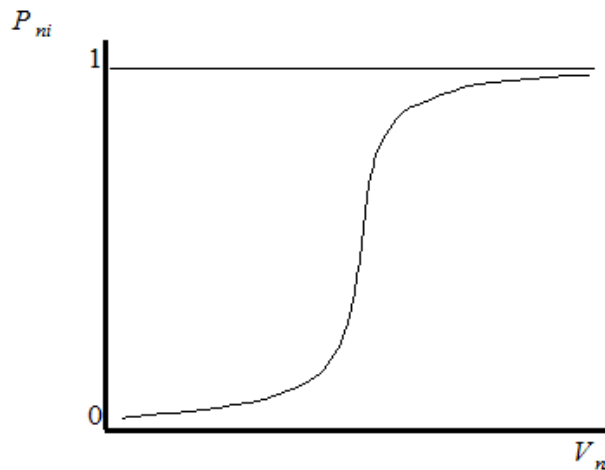


Figura 10. Relação entre a probabilidade determinada e a componente determinística da utilidade (Train 2009).

O *rank-ordered logit* é derivado a partir do modelo *logit* padrão, podendo ser utilizado quando a variável resposta é uma ordenação de alternativas (*ranking*). Considerando os pressupostos do modelo *logit* padrão, a probabilidade de observar uma determinada ordenação pode ser determinada através da multiplicação das expressões *logit*. Perante um conjunto de alternativas, 1, 2, 3, 4 em que os erros ε_{nj} são independentes, distribuídos identicamente e possuem uma distribuição de valor extremo, se o decisor prefere a seguinte ordenação $2 > 4 > 3 > 1$, atendendo à definição do modelo *logit* considera-se que o decisor escolhe a alternativa

2 como sendo a melhor de entre o conjunto de alternativas 1, 2, 3, 4, escolhe a alternativa 4 como a melhor do conjunto de alternativas 1, 3 e assim sucessivamente.

Nesta situação a probabilidade de obter a ordenação $2 > 4 > 3 > 1$ é dada pela seguinte expressão:

$$P(2 > 4 > 3 > 1) = \frac{e^{\beta \cdot x_{n2}}}{\sum_{j=1,2,3,4} e^{\beta \cdot x_{nj}}} \frac{e^{\beta \cdot x_{n4}}}{\sum_{j=1,3,4} e^{\beta \cdot x_{nj}}} \frac{e^{\beta \cdot x_{n3}}}{\sum_{j=3,1} e^{\beta \cdot x_{nj}}} \quad (77)$$

O modelo é definido pelo vetor β que constitui o conjunto de parâmetros lineares que multiplicam os atributos observáveis. Este vetor é calculado através da maximização da função de verosimilhança. Para um conjunto de decisores, N , o logaritmo da função de verosimilhança é dado por:

$$\log l(\beta) = \sum_{n=1}^N \sum_{j=1}^J V_{nj} - \sum_{n=1}^N \sum_{j=1}^J \log \left[\sum_{k=j}^J \delta_{nk} e^{V_{nj}} \right] \quad (78)$$

McFadden 1974 mostrou que esta função é côncava em β e, portanto, possui um máximo. A maximização desta função providencia os parâmetros do modelo. O valor obtido é o logaritmo da probabilidade de se observar uma dada resposta utilizando este modelo.

Independência das Alternativas Irrelevantes

Estes modelos pressupõem que se verifica uma propriedade designada por “Independência das Alternativas Irrelevantes” (*IIA*). A *IIA* significa que a razão das probabilidades de duas alternativas é sempre a mesma independentemente da existência de alternativas adicionais. É neste sentido que se considera aquela razão independente de alternativas irrelevantes. Isto é facilmente demonstrado por:

$$\frac{P_{ni}}{P_{nk}} = \frac{\frac{e^{\beta \cdot x_{ni}}}{\sum_j e^{\beta \cdot x_{nj}}}}{\frac{e^{\beta \cdot x_{nk}}}{\sum_j e^{\beta \cdot x_{nj}}}} = \frac{e^{\beta \cdot x_{ni}}}{e^{\beta \cdot x_{nk}}} \quad (79)$$

A esta propriedade estão associadas as seguintes vantagens:

- permite estimar os parâmetros do modelo considerando um subconjunto de alternativas de entre o total de alternativas disponíveis para o decisor. Isto é particularmente importante numa situação em que o conjunto de alternativas possui uma dimensão considerável e a estimação dos parâmetros do modelo torna-se morosa devido a limitações computacionais;
- se o analista apenas está interessado em analisar as escolhas efetuadas pelos decisores relativas a um subconjunto de alternativas a propriedade *IIA* permite a estimação dos parâmetros do modelo excluindo as restantes alternativas. Este procedimento permite

ao analista reduzir o tempo necessário e o custo associados à recolha de dados relativamente às alternativas excluídas.

Para ser possível aplicar o modelo *logit* e o *rank-ordered logit* deverá ser satisfeita a propriedade *IIA*. Para verificar esta propriedade utilizou-se o teste de *Hausman* (Hausman, J. e McFadden 1984; Hausman, J. A. 1978). Neste teste determinam-se os parâmetros do modelo para duas situações. Em primeiro lugar os parâmetros são estimados utilizando um subconjunto de alternativas do total de alternativas disponíveis. Em segundo lugar os parâmetros são calculados utilizando-se o conjunto total de alternativas disponíveis. Se a propriedade *IIA* for satisfeita então os parâmetros estimados das duas formas não deverão diferir significativamente entre si.

Robustez do modelo

A amostra utilizada é constituída por um conjunto de investigadores que trabalham em áreas científicas com culturas de publicação e citação próprias e que diferem entre si. Por outro lado, o número de concursos usado não é de grande dimensão e o peso atribuído ao desempenho científico varia de concurso para concurso. Isto poderá originar um modelo pouco robusto quando são introduzidas pequenas alterações.

Assim, para estudar a robustez e estabilidade de cada um dos modelos foram utilizadas técnicas de reamostragem. Estas técnicas baseiam-se no cálculo de estimativas a partir de um conjunto de subamostras formadas a partir da amostra inicial.

A técnica de reamostragem utilizada foi a designada por *jackknife*. Esta metodologia é aplicada com frequência para estimar a variância ou a tendência de um estimador. Utiliza-se um processo iterativo em que em cada iteração é retirada uma observação do conjunto de dados iniciais e os parâmetros em questão são estimados utilizando as restantes observações. O processo passa por tantas iterações quanto o número de observações na amostra.

O procedimento adotado neste estudo foi o seguinte:

- a) o processo envolveu 27 iterações. Um concurso foi removido em cada iteração e os parâmetros dos modelos foram determinados utilizando os restantes concursos. Os parâmetros estimados foram posteriormente utilizados no cálculo da probabilidade de cada um dos candidatos do concurso omitido ser selecionado em primeiro lugar;
- b) terminado o processo iterativo o coeficiente de correlação de *Pearson* foi calculado para estudar a relação entre as probabilidades determinadas por este método e as calculadas pelos modelos estimados com a amostra completa.

Com este procedimento foi possível inferir sobre as alterações provocadas nas probabilidades determinadas para cada candidato e concluir sobre a robustez do modelo.

Derivadas e elasticidades

A utilidade determinada para cada indivíduo depende de um conjunto de atributos relacionados com as alternativas e que são observáveis pelo analista. É assim importante

analisar a influência destes atributos na utilidade. Isto pode ser efetuado através da determinação de um conjunto de elasticidades.

As elasticidades foram determinadas para as seguintes situações:

- ***Alteração da utilidade (V_{nj}) quando os atributos de x_{nj} são alterados***

Nesta situação é avaliado em quanto o valor absoluto da utilidade é incrementado quando se aumenta em uma unidade o valor dos atributos que compõem x_{nj} . O que se pretende nesta situação é a derivada parcial da utilidade em relação a cada uma das variáveis observáveis:

$$\frac{\partial V_{nj}}{\partial x_{nj}} (80)$$

- ***Alteração relativa de V_{nj} devido a uma alteração relativa dos atributos de x_{nj}***

Aqui o que se pretende é determinar a alteração proporcional em V_{nj} para uma alteração proporcional dos atributos que representam x_{nj} .

$$\frac{\partial V_{nj}}{\partial x_{nj}} \frac{x_{nj}}{V_{nj}} (81)$$

- ***Alteração relativa de V_{nj} quando os atributos de x_{nj} são alterados***

Nesta situação é analisada a variação proporcional em V_{nj} para uma alteração do valor absoluto dos atributos que constituem x_{nj} .

$$\frac{\partial V_{nj}}{\partial x_{nj}} \frac{1}{V_{nj}} (82)$$

- ***Alteração em V_{nj} devido a uma alteração relativa dos atributos que compõem x_{nj}***

O incremento no valor absoluto de V_{nj} é avaliado quando existe uma alteração proporcional dos atributos de x_{nj} .

$$\frac{\partial V_{nj}}{\partial x_{nj}} x_{nj} (83)$$

Todas as situações que envolveram análises estatísticas foram realizadas recorrendo aos softwares *IBM SPSS Statistics* e o *Stata/IC12*.

4.3 Definição do modelo

A presença de multicolineariedade entre as variáveis independentes é um fenómeno que ocorre com alguma frequência num conjunto de dados. Quando isto se verifica torna-se difícil determinar qual a variável independente que melhor explica as variações da variável dependente. Na presença de multicolineariedade os coeficientes estimados podem sofrer grandes alterações quando são introduzidas pequenas modificações.

Os valores do “fator de inflação da variância” (*VIF*) são apresentados na Tabela 37. O valor do “índice de condição” (*CI*) foi também calculado.

Os resultados mostram que os indicadores *NDF* e h_{nf} são os que possuem maior *VIF*. O valor 10 é sugerido na literatura como valor limite (Pestana e Gageiro 1998). O indicador *NDF* possui um *VIF* que não está distante do valor limite e a variável h_{nf} possui um valor superior sugerindo assim que está presente a multicolineariedade. O valor obtido para o *CI* foi de 8,12, inferior ao valor limite recomendado, 15 (Pestana e Gageiro 1998).

Tabela 37. Valores obtidos para o *VIF*.

Variável	<i>NDF</i>	<i>HCD</i>	<i>CD</i>	<i>PDC</i>	h_{nf}	<i>NIR</i>	<i>SNIP_m</i>	<i>SJR_m</i>	<i>NI</i>	<i>QI</i>	<i>DIC</i>	<i>NA_m</i>
<i>VIF</i>	8,09	2,47	1,17	2,10	10,62	3,06	1,51	1,28	2,19	2,08	1,26	1,39

Uma vez que os resultados apontam para a presença de multicolineariedade três estratégias, que levaram à definição não de um modelo, mas sim de três, foram adotadas com o objetivo de a eliminar.

- Estratégia 1 – agrupar o conjunto de 12 indicadores utilizando a análise fatorial.
- Estratégia 2 – eliminar o indicador h_{nf} e prosseguir com a definição do modelo utilizando os 11 indicadores restantes.
- Estratégia 3 - eliminar o indicador *NDF* e prosseguir com a definição do modelo utilizando os 11 indicadores restantes.

4.3.1 Análise Fatorial

A análise de observações atípicas, utilizando as metodologias mencionadas na secção deste capítulo referente à descrição da metodologia, resultou na eliminação de três observações (atípicas) da amostra considerada. As estatísticas descritivas para a amostra são apresentadas na Tabela 38 em relação a cada um dos indicadores (variáveis).

Após a remoção das observações atípicas procedeu-se à análise da normalidade. Na realidade o método utilizado na extração dos fatores (componentes principais) não exige que as variáveis apresentem uma distribuição Normal. No entanto optou-se por analisar este aspeto e nos casos em que uma transformação da variável resultou numa distribuição Normal a mesma foi adotada.

Os resultados do teste de *Shapiro-Wilk* são apresentados na Tabela 39 para cada um dos indicadores (variáveis) originais.

No teste de *Shapiro-Wilk* são testadas as seguintes hipóteses:

H_0 : A amostra provém de uma distribuição Normal.

H_1 : A amostra não provém de uma distribuição Normal (teste bilateral).

Os resultados da Tabela 39 mostram que apenas as variáveis *NDF*, *NIR*, *h_{nf}* e *SNIP_m* seguem uma distribuição Normal ($p > 0,05$).

Tabela 38. Estatísticas descritivas para a amostra de investigadores.

Variável	Média	Mediana	Desvio padrão	Coefficiente de variação (%)	Máximo	Mínimo	Simetria
<i>NDF</i>	10,84	8,78	7,79	71,86	36,71	0,45	1,345
<i>h_{nf}</i>	3,42	3,09	2,01	58,77	8,56	0,20	0,615
<i>NIR</i>	4,45	3,76	2,82	63,37	14,44	0,23	1,328
<i>HCD</i>	20	18	13	65	66	0	0,612
<i>SNIP_m</i>	1,049	1,010	0,347	0,033	2,135	0,21	0,292
<i>SJR_m</i>	0,165	0,120	0,152	92,121	0,891	0,039	2,663
<i>DIC</i>	44	40	26	59	96	0	0,212
<i>CD</i>	91	93	6	7	100	61	-2,129
<i>NI</i>	1,2	1,2	0,1	8,3	1,6	0,8	0,560
<i>QI</i>	51,1	50,8	4,3	8,4	68,8	38,4	0,409
<i>NA_m</i>	6,22	6,47	1,78	28,62	11,5	0,48	-0,754
<i>PDC</i>	62	66	17	27,42	100	14	-0,868

Tabela 39. Valores obtidos no teste de *Shapiro-Wilk*.

Variável	Shapiro-Wilk		
	Estatística	df	p valor
log(<i>NDF</i>)	0,99	171	0,29
log(<i>NIR</i>)	0,99	171	0,09
sqrt(<i>h_{nf}</i>)	0,99	171	0,13
<i>HCD</i>	0,96	171	0,00
<i>NI</i>	0,93	171	0,00
<i>QI</i>	0,95	171	0,00
<i>PDC</i>	0,95	171	0,00
<i>SJR_m</i>	0,69	171	0,00
<i>SNIP_m</i>	0,99	171	0,24
<i>DIC</i>	0,97	171	0,00
<i>CD</i>	0,83	171	0,00
<i>NA_m</i>	0,95	171	0,00

A análise fatorial parte da matriz de correlações para chegar à estrutura fatorial. Torna-se essencial que um número considerável de correlações estejam presentes para a análise fatorial ser viável. Na Tabela 40 estão apresentadas as correlações entre os indicadores.

Tabela 40. Matriz de correlações para os 12 indicadores (variáveis) utilizados.

	<i>NDF</i>	<i>NIR</i>	<i>h_{nf}</i>	<i>HCD</i>	<i>NI</i>	<i>QI</i>	<i>PDC</i>	<i>SJR_m</i>	<i>SNIP_m</i>	<i>DIC</i>	<i>CD</i>	<i>NA_m</i>
<i>NDF</i>	1,00	0,50	0,92	0,33	0,23	0,17	0,39	0,04	0,17	0,00	-0,14	-0,22
<i>NIR</i>	-	1,00	0,71	0,66	0,15	0,06	0,67	0,21	0,41	0,21	0,10	0,17
<i>h_{nf}</i>	-	-	1,00	0,48	0,22	0,11	-0,60	0,13	0,30	0,09	-0,06	-0,07
<i>HCD</i>	-	-	-	1,00	0,14	0,15	0,43	0,25	0,13	0,11	-0,03	0,22
<i>NI</i>	-	-	-	-	1,00	0,66	0,11	0,03	0,10	0,17	0,00	-0,16
<i>QI</i>	-	-	-	-	-	1,00	-0,06	-0,05	-0,02	0,13	-0,04	-0,09
<i>PDC</i>	-	-	-	-	-	-	1,00	0,35	0,43	0,28	0,08	0,13
<i>SJR_m</i>	-	-	-	-	-	-	-	1,00	0,26	0,20	0,22	0,05
<i>SNIP_m</i>	-	-	-	-	-	-	-	-	1,00	0,15	0,20	-0,11
<i>DIC</i>	-	-	-	-	-	-	-	-	-	1,00	0,12	0,11
<i>CD</i>	-	-	-	-	-	-	-	-	-	-	1,00	0,05
<i>NA_m</i>	-	-	-	-	-	-	-	-	-	-	-	1,00

Para todas as correlações superiores a 0,3 verificou-se se as mesmas eram significativas. No total cerca de 23% das correlações são significativas ($p < 0,001$). As variáveis *DIC*, *CD* e *NA_m* são as únicas que não apresentam nenhuma correlação igual ou superior a 0,3 com outras variáveis. Por outro lado as variáveis *NDF* e *h_{nf}* estão fortemente correlacionadas. Como mencionado no Capítulo 3 o *h_{nf}* é influenciado pelo número de documentos e portanto esta correlação já era esperada.

Na análise fatorial elevadas correlações não são apropriadas (Hair 2010). No entanto, optou por manter-se ambas as variáveis por se considerar que o *NDF* é uma variável importante quando se pretende descrever o desempenho científico de um investigador com base na dimensão quantidade. Por outro lado, o *h_{nf}* apresenta propriedades que lhe permitem conciliar uma medida da quantidade e do impacto da produção científica de um investigador. Este indicador também apresenta um conjunto de propriedades que o tornam atrativo na descrição e comparação do desempenho científico de indivíduos (ver Capítulo 3).

As correlações parciais também foram analisadas. Correlações parciais muito elevadas não devem estar presentes. Na matriz anti-imagem as correlações parciais são os valores simétricos dos valores que se encontram fora da diagonal. Na Tabela 41 é apresentada a matriz anti-imagem. Os valores na diagonal representam a medida de adequação da amostra (*MSA*).

Tabela 41. Matriz anti-imagem.

	<i>NDF</i>	<i>NIR</i>	<i>h_{nf}</i>	<i>HCD</i>	<i>NI</i>	<i>QI</i>	<i>PDC</i>	<i>SJR_m</i>	<i>SNIP_m</i>	<i>DIC</i>	<i>CD</i>	<i>NA_m</i>
<i>NDF</i>	0,57	0,25	-0,93	0,13	0,04	-0,13	-0,29	-0,02	0,17	0,03	0,03	0,27
<i>NIR</i>	-	0,79	-0,39	-0,45	0,02	-0,01	0,17	0,13	-0,20	-0,08	-0,14	-0,09
<i>h_{nf}</i>	-	-	0,63	-0,13	-0,07	0,11	0,37	0,04	-0,15	0,02	-0,03	-0,16
<i>HCD</i>	-	-	-	0,75	0,02	-0,15	-0,06	-0,23	0,19	0,08	0,13	-0,10
<i>NI</i>	-	-	-	-	0,59	-0,65	0,07	-0,001	-0,02	-0,08	-0,03	0,13
<i>QI</i>	-	-	-	-	-	0,52	-0,14	0,04	-0,01	-0,10	0,01	-0,05
<i>PDC</i>	-	-	-	-	-	-	0,80	0,24	0,13	0,18	-0,08	0,05
<i>SJR_m</i>	-	-	-	-	-	-	-	0,67	-0,14	-0,10	-0,19	0,03
<i>SNIP_m</i>	-	-	-	-	-	-	-	-	0,74	0,01	-0,11	0,23
<i>DIC</i>	-	-	-	-	-	-	-	-	-	0,78	-0,05	-0,06
<i>CD</i>	-	-	-	-	-	-	-	-	-	-	0,59	0,06
<i>NA_m</i>	-	-	-	-	-	-	-	-	-	-	-	0,52

As correlações parciais são no geral baixas, o que está de acordo com o sugerido por Hair 2010. A variável NDF e h_{nf} apresentam uma correlação parcial elevada, comportamento que já era esperado pelo mencionado anteriormente.

Apesar de ser possível observar que existem indicadores que apresentam alguma correlação com outros indicadores, isto não é suficiente para garantir a validade da aplicação da análise fatorial. A realização do teste de *Bartlett*, da estatística de *Kaiser-Meyer-Olkin* (*KMO*) e a análise da matriz anti-imagem foram utilizados para testar a viabilidade da análise fatorial perante o conjunto de dados.

O teste de esfericidade de *Bartlett* testa as seguintes hipóteses:

H_0 : A matriz de correlações é uma matriz identidade.

H_1 : A matriz de correlações não é a matriz identidade (teste bilateral).

Os resultados obtidos para o teste levam à rejeição da hipótese nula ($p < 0,001$), isto é, há evidência estatística de que existe correlação significativa entre as variáveis. O valor obtido para o *KMO* foi de 0,671 e a matriz anti-imagem mostra que todas as variáveis possuem um *MSA* superior a 0,5.

Consoante o valor obtido para o *KMO* as seguintes conclusões podem ser retiradas sobre a aplicabilidade da análise fatorial ao conjunto de dados (Maroco 2007):

- valores de *KMO* entre]0,9-1,0] – excelente;
- valores de *KMO* entre]0,8-0,9] – boa;
- valores de *KMO* entre]0,7-0,8] – média;
- valores de *KMO* entre]0,6-0,7] – medíocre;
- valores de *KMO* entre]0,5-0,6] – mau mas ainda aceitável;
- valores de *KMO* entre $\leq 0,5$ – inaceitável.

Considerando o valor obtido para o *KMO*, que o valor de *MSA* é superior a 0,5 como recomendado (Maroco 2007) e o resultado do teste de *Bartlett*, prosseguiu-se com a aplicação da análise fatorial.

A aplicação da análise fatorial aos dados iniciais mostrou que o indicador NA_m fica isolado em um único fator, isto é, o indicador apenas possui um peso fatorial significativo em um fator. De facto, apresenta uma baixa correlação com os restantes indicadores o que contribui para o resultado observado. Perante isto optou-se pela eliminação do indicador da análise.

Após a eliminação do NA_m a análise fatorial foi novamente aplicada aos restantes dados.

Uma vez que a análise fatorial agrupa as variáveis que estão correlacionadas entre si é importante considerar a variância de uma variável que é partilhada pelas restantes variáveis. Esta variância é denominada por comunalidade. O valor mínimo estabelecido para a comunalidade foi de 0,5 (Hair 2010). Da aplicação da análise fatorial resultou que a comunalidade do $SNIP_m$ e DIC não obedece ao critério estabelecido. Perante estes resultados as variáveis foram eliminadas. Do total de 12 variáveis iniciais (indicadores) foram eliminadas três.

Perante o conjunto das nove variáveis restantes a análise fatorial foi aplicada novamente. O *KMO* tem agora o valor de 0,64, da aplicação do teste de *Bartlett* resulta que a matriz de correlações não é a matriz identidade e todos os indicadores possuem um *MAS* igual ou superior a 0,5. Na Tabela 42 é apresentada a variância total explicada por cada fator e pelo conjunto de fatores que possuem um valor próprio superior a um.

Tabela 42. Proporção da variância total explicada por cada fator após a rotação.

Fator	Valor próprio inicial	Proporção da variância explicada (%)	Proporção da variância explicada acumulada (%)
1	3,469	37,193	37,193
2	1,668	18,899	56,091
3	1,286	15,268	71,360
4	0,813	-	-
5	0,702	-	-
6	0,514	-	-
7	0,300	-	-
8	0,215	-	-
9	0,033	-	-

Segundo o critério de *Kaiser* (Kaiser 1960) apenas os fatores com valores próprios superiores a um devem ser retidos pois consideram uma proporção da variância significativa. Aplicando este critério, três fatores foram retidos, explicando 71,4% do total da variância dos dados. Embora o valor alcançado não seja elevado, é importante ter em consideração que os indicadores selecionados pretendem descrever o desempenho científico que é algo complexo, multidimensional e que é influenciado por um grande conjunto de fatores.

Na Tabela 43 estão representados os pesos fatoriais (após a rotação) e as comunalidades obtidas após a extração.

Tabela 43. Pesos fatoriais e comunalidades obtidos na análise fatorial.

Variável	Pesos fatoriais			Comunalidade
	Fator 1	Fator 2	Fator 3	
h_{nf}	0,929	-	-	0,901
<i>NIR</i>	0,855	-	-	0,772
<i>NDF</i>	0,796	-	-0,333	0,773
<i>PDC</i>	0,758	-	-0,300	0,672
<i>HCD</i>	0,675	-	-	0,511
<i>QI</i>	-	0,917	-	0,844
<i>NI</i>	-	0,893	-	0,818
<i>CD</i>	-	-	0,739	0,553
<i>SJR_m</i>	-	-	0,718	0,579

O fator 1 representa 37,2% do total da variância explicada pelos três fatores. O h_{nf} , *NIR*, *NDF*, *PDC* e *HCD* são os indicadores que possuem os maiores pesos fatoriais neste fator. O fator 2 explica 18,9% do total da variância e os indicadores com os pesos fatoriais mais

elevados são o *NI* e o *QI*. O fator 3 explica uma menor fração do total da variância, 15,3%, e os indicadores *CD* e *SJR_m* são os que apresentam maior correlação com este fator.

Em Hair 2010 é discutida a dimensão que uma amostra deve possuir para inferir sobre a significância dos pesos fatoriais. Para uma amostra cuja dimensão varia entre [150-200] observações os pesos fatoriais apenas são considerados significativos se o valor for no mínimo igual a 0,45 (dimensão 150) e igual a 0,40 (dimensão 200). Nos resultados obtidos é possível constatar que os indicadores com maior contribuição para a definição de cada fator apresentam todos pesos fatoriais significativos.

Para todos os indicadores, no mínimo, 50% da variância (comunalidades superiores a 0,5) de cada variável é explicada pelos três fatores identificados.

Com a estrutura fatorial definida é importante avaliar a consistência interna ou a confiabilidade da estrutura obtida. Isto permite verificar se os indicadores individuais ou conjunto de indicadores originam resultados consistentes. O *Cronbach's α* foi utilizado para este fim. Valores de *Cronbach's α* compreendidos entre [0,7-0,8] são aceitáveis (Field 2005). O *Cronbach's α* é apresentado para cada fator na Tabela 44.

Tabela 44. Valores obtidos para o *Cronbach's α* .

Fator	<i>Cronbach's α</i>
1	0,7
2	0,8
3	0,4

O valor de *Cronbach's α* obtido para o fator 3 não se encontra dentro do intervalo recomendado. A heterogeneidade deste fator poderá explicar o resultado obtido. Estes são indicadores que estão de alguma forma relacionados com a comunidade científica em que se insere o candidato. O indicador *CD* representa a comunidade que de alguma forma utiliza o trabalho efetuado pelo investigador para a ajudar a resolver o seu próprio trabalho. No *SJR_m* a comunidade científica em que este se insere será à partida mais heterogênea do que aquela que se encontra em *CD* e portanto valores muito elevados para o *Cronbach's α* não podem ser esperados para o fator 3.

Aos fatores obtidos podem ser atribuídas designações de acordo com o que os indicadores que os definem representam:

- os indicadores com os maiores pesos fatoriais no fator 1 estão relacionados com a contribuição de cada investigador para o conhecimento científico. Esta dimensão pode ser designada por “Contribuição Pessoal” (*PC*);
- os indicadores mais importantes no fator 2 estão relacionados com a visibilidade das instituições onde foram desenvolvidas as atividades científicas do investigador. A dimensão pode ser intitulada como “Prestígio da Instituição” (*PI*);
- no fator 3 os indicadores com maior peso fatorial estão de alguma forma relacionados com a comunidade científica que partilha os mesmos interesses científicos do investigador. A esta dimensão foi atribuída a designação “Comunidade Científica” (*SC*).

Finalizada a aplicação da análise fatorial ao conjunto de dados disponíveis a multicolineariedade foi novamente avaliada agora para cada uma das estratégias mencionadas anteriormente. Os resultados encontram-se na Tabela 45. As variáveis compostas (*PC*; *PI*; *SC*) que resultaram da análise fatorial foram utilizadas em conjunto com as variáveis excluídas e um modelo definido a partir destas.

Tabela 45. Valores de *VIF* para as variáveis e do *CI*.

Índices	Variáveis	Estratégia 1	Estratégia 2	Estratégia 3
VIF	<i>PC</i>	1,17	-	-
	<i>PI</i>	1,05	-	-
	<i>SC</i>	1,24	-	-
	<i>NDF</i>	-	1,49	-
	<i>HCD</i>	-	2,40	2,43
	<i>CD</i>	-	1,14	1,15
	<i>PDC</i>	-	2,00	2,09
	<i>h_{nf}</i>	-	-	1,96
	<i>NIR</i>	-	2,69	2,88
	<i>SNIP_m</i>	1,31	1,46	1,45
	<i>SJR_m</i>	-	1,28	1,28
	<i>NI</i>	-	2,19	2,18
	<i>QI</i>	-	2,05	2,04
	<i>DIC</i>	1,14	1,25	1,23
	<i>NA_m</i>	1,12	1,34	1,25
	<i>CI</i>	1,89	1,75	3,83

Os resultados obtidos mostram que os problemas de multicolineariedade foram eliminados. Perante a ausência de multicolineariedade foi possível prosseguir para a definição dos modelos utilizando como metodologia o *ROLR*.

4.3.2 Rank-ordered logit regression

A aplicação desta metodologia apenas é viável se a independência das alternativas irrelevantes (*IIA*) for respeitada. A *IIA* foi verificada utilizando o teste de *Hausman*. As seguintes hipóteses foram testadas:

H_0 : Os parâmetros estimados para o modelo utilizando um subconjunto de alternativas não são diferentes dos estimados quando todas as alternativas são consideradas.

H_1 : Os parâmetros estimados para o modelo utilizando um subconjunto de alternativas são diferentes dos estimados quando todas as alternativas são consideradas (teste bilateral).

Para testar a hipótese *IIA* utilizou-se o conjunto de dados iniciais (isto é, as ordenações completas) e os coeficientes obtidos foram comparados com os coeficientes determinados para um conjunto de dados em que foram eliminados alguns dos candidatos (alternativas). As variáveis com impacto significativo quando são utilizados subconjuntos de alternativas

diferem daquelas que são obtidas quando se recorre ao conjunto completo de dados. As seguintes razões poderão explicar este comportamento:

- é razoável considerar, que no processo de seleção, os avaliadores (pares) colocam maior esforço na ordenação dos melhores candidatos do que nos candidatos que à partida serão colocados nas posições baixas. No final, apenas um ou dois candidatos serão aceites. A seriação dos candidatos nas últimas posições das ordenações é irrelevante para as universidades. Isto sugere que maior peso deve ser atribuído às posições de topo nas seriações finais;
- o número de candidatos varia muito de concurso para concurso. Apenas cinco concursos têm dez candidatos. Como o número de concursos com um grande número de candidatos é pequeno, tal poderá induzir a uma perda de eficiência da metodologia utilizada. A distribuição do número de candidatos nos 27 concursos é mostrada na Tabela 46.

Tabela 46. Distribuição dos candidatos nos 27 concursos.

Número de candidatos	Número de concursos
1 ou mais	27
2 ou mais	27
3 ou mais	21
4 ou mais	19
5 ou mais	18
6 ou mais	17
7 ou mais	14
8 ou mais	12
9 ou mais	10
10 ou mais	5
Apenas 11	1

Perante estes factos, optou-se por alterar a variável dependente (ordenação final):

- para cada ordenação ao candidato na primeira posição foi atribuído o número 1 e aos restantes candidatos foi atribuído zero. Isto é o mesmo que pedir ao júri que selecione o candidato que considera superior a todos os outros. Neste caso a variável dependente foi designada por B1;
- para cada ordenação, aos candidatos na primeira e segunda posição foi atribuído o número 1 e 2 e aos candidatos nas restantes posições o número zero. À variável dependente foi denotada por B2.

Este processo foi repetido adicionando de cada vez uma nova posição obtendo-se no final as variáveis B1, B2, ... até B11. Para cada uma destas variáveis dependentes e para as três estratégias definidas foi aplicado o *ROLR*. Os resultados obtidos são apresentados na Tabela 47.

Tabela 47. Variáveis com impacto significativo ($p < 0,05$) retidas quando são utilizadas as diferentes variáveis dependentes.

Variável dependente	Estratégia 1	Estratégia 2	Estratégia 3
B1	<i>PC</i>	h_{nf}	<i>NDF</i>
B2	<i>PC</i>	h_{nf}	<i>NDF, HCD</i>
B3	<i>PC</i>	h_{nf}, HCD	<i>NDF, HCD</i>
B4	<i>PC</i>	h_{nf}, HCD	<i>NDF, HCD, NA_m</i>
B5	<i>PC</i>	h_{nf}	<i>NDF, HCD, NA_m</i>
B6	<i>PC</i>	h_{nf}, NA_m	<i>NDF, HCD, NA_m</i>
B7	<i>PC</i>	h_{nf}, NA_m	<i>NDF, HCD, NA_m</i>
B8	<i>PC</i>	h_{nf}, NA_m, SJR_m	<i>NDF, NA_m, SJR_m</i>
B9	<i>PC</i>	h_{nf}, NA_m, SJR_m	<i>NDF, HCD, NA_m, SJR_m, NI, Q1</i>
B10	<i>PC</i>	h_{nf}, NA_m, SJR_m	<i>NDF, HCD, NA_m, SJR_m</i>
B11	<i>PC</i>	h_{nf}, NA_m, SJR_m	<i>NDF, HCD, NA_m, SJR_m, NI, Q1</i>

Para a estratégia 1 verificou-se que apenas a variável *PC* é retida em todos os casos. Na estratégia 2 apenas a variável h_{nf} possui impacto significativo em todas as variáveis dependentes e na estratégia 3 o mesmo é observado para a variável *NDF*.

Chapman e Staelin 1982 sugerem o uso do índice de razão da máxima verosimilhança (ρ) para decidir quais as posições das ordenações de alternativas que devem ser utilizadas na definição de um modelo baseado no *ROLR*. Os autores mencionam que quanto maior o número de posições consideradas de uma dada ordenação, maior o número de observações independentes e isto contribuirá, em princípio, para diminuir o desvio padrão dos parâmetros estimados. No entanto, os mesmos consideram que a utilização das posições mais baixas quando um grande número de alternativas está disponível poderá introduzir algum viés nos parâmetros estimados. Os autores mencionam que estas posições são, no mínimo, menos sistemáticas no que respeita às regras utilizadas na escolha das alternativas colocadas nas posições de topo. Para os candidatos colocados nas posições mais baixas das ordenações os avaliadores poderão não ser capazes de identificar o valor relativo das utilidades e portanto a ordenação destes candidatos poderá incluir alguma aleatoriedade. O índice ρ foi aqui adotado para a determinação da variável dependente mais adequada.

$$\rho = 1 - \frac{\log l(\beta)}{\log l(0)} \quad (84)$$

O $l(\beta)$ é a função da máxima verosimilhança para o modelo com os parâmetros estimados e o $l(0)$ é a função da máxima verosimilhança do modelo nulo (isto é sem parâmetros). O valor deste índice varia entre 0 e 1 e valores próximos de 1 são desejáveis (Train 2009 Capítulo 3). Os resultados obtidos são apresentados na Figura 11.

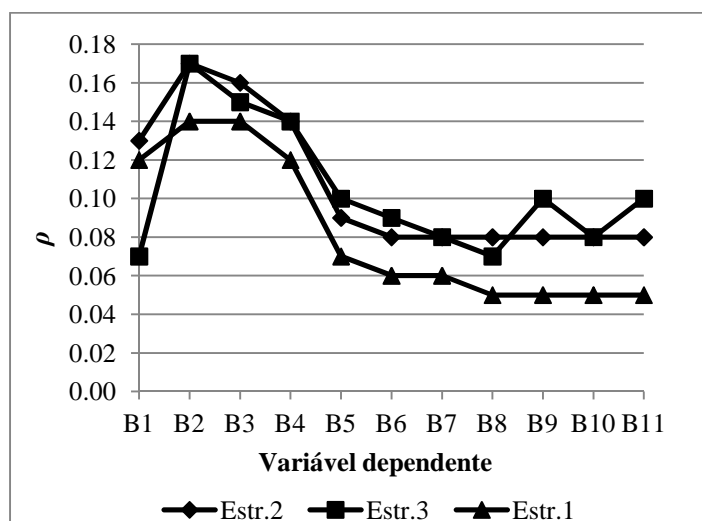


Figura 11. Índice da razão de máxima verosimilhança para as diferentes variáveis dependentes.

Quando a um indivíduo é pedido para efetuar a ordenação de um conjunto de alternativas de acordo com a utilidade que atribui a cada uma delas, em teoria, é esperado que a estimação dos parâmetros do modelo seja mais eficiente do que ocorre numa situação onde apenas lhe é pedido para seleccionar a alternativa preferida (Chapman e Staelin 1982; Fok, Paap e Van Dijk 2012). Isto poderá explicar o aumento do ρ quando se passa da variável B1 para a variável B2. O ρ diminui, lentamente, até B4 e permanece quase estável desde B5 até B11 apesar de o número de variáveis com impacto significativo para algumas destas variáveis (B5-B11) ser superior. Para a variável B4 o número de concursos diminui consideravelmente relativamente a B2 como se pode constatar na Tabela 46. O verificado poderá justificar o comportamento de ρ em B4 relativamente a B2.

A diminuição acentuada do valor de ρ de B4 para B5 sugere que na definição dos modelos apenas as primeiras 4 posições de cada ordenação sejam consideradas atribuindo uma posição ex-aequo aos candidatos colocados nas posições mais baixas. Nesta situação a diminuição do número de concursos não parece ser a razão que justifica o comportamento observado pois a diferença é apenas de um concurso. Os resultados obtidos vão de encontro ao mencionado por Chapman e Staelin 1982, isto é, os candidatos colocados nas posições mais baixas poderão não ter sido posicionados com o mesmo rigor com que foram aqueles colocados nas posições cimeiras. Daí a opção por se considerarem apenas as 4 posições iniciais de cada ordenação. Uma posição ex-aequo foi considerada para os candidatos que ocupavam uma posição inferior a 4 em cada ordenação pois considera-se que existe alguma informação disponível, isto é, estes candidatos possuem associada uma utilidade inferior (para cada júri) aos restantes candidatos.

Após a seleção da variável dependente mais adequada o teste de *Hausman* foi realizado novamente. O resultado do teste permitiu concluir que a hipótese *IIA* é verificada ($p > 0,05$).

A aplicação do *ROLR* ao conjunto de dados disponíveis permitiu a definição de três modelos. Os parâmetros estimados para cada modelo são apresentados na Tabela 48.

Tabela 48. Parâmetros estimados para cada modelo.

Modelo	Variável independente	Coefficiente	Desvio padrão	Z	P> z	LR chi2	P>chi2
1	PC	1,06	0,21	5,06	<0,001*	33,38	<0,001*
2	h_{nf}	0,40	0,09	4,61	<0,001*	40,61	<0,001*
	HCD	0,032	0,01	2,45	0,014*		
3	NDF	0,11	0,02	4,58	<0,001*	40,99	<0,001*
	HCD	0,044	0,01	3,47	0,001*		
	NA _m	0,19	0,09	2,07	0,038*		

Z – estatística do teste

*Significativo ao nível de 0,05.

LR chi2 – *likelihood ratio chi-square test*.

Os resultados do teste de LRchi2, cujas hipóteses de teste são:

H_0 : Todos os coeficientes de regressão do modelo são iguais a zero.

H_1 : No mínimo um dos coeficientes de regressão do modelo não é zero (teste bilateral).

conduzem à rejeição da hipótese nula, $p<0,001$.

Os indicadores apresentados na Tabela 48 são aqueles que de entre o conjunto de indicadores utilizados têm impacto significativo ($p<0,05$). Todos os indicadores com impacto significativo têm um efeito positivo na probabilidade de um dado candidato ser selecionado em primeiro lugar, isto é, a probabilidade aumenta com o incremento do valor destes indicadores.

Com os parâmetros estimados é possível proceder ao cálculo da probabilidade de um candidato ser selecionado em primeiro lugar utilizando as seguintes expressões:

Modelo 1

$$P_{ni} = \frac{e^{1,06PC_{ni}}}{\sum_j e^{1,06PC_{nj}}} \quad (85)$$

Modelo 2

$$P_{ni} = \frac{e^{0,40h_{nf_{ni}}+0,032HCD_{ni}}}{\sum_j e^{0,40h_{nf_{nj}}+0,032HCD_{nj}}} \quad (86)$$

Modelo 3

$$P_{ni} = \frac{e^{0,11NDF_{ni}+0,044HCD_{ni}+0,19NA_{mni}}}{\sum_j e^{0,11NDF_{nj}+0,044HCD_{nj}+0,19NA_{mni}}} \quad (87)$$

Uma vez definidos os modelos prossegue-se o estudo de algumas das características que estes apresentam.

Robustez dos modelos

A metodologia de reamostragem *jackknife* foi utilizada para avaliar a robustez de cada modelo. Na Figura 12 é apresentada a relação entre a probabilidade de um candidato ficar em primeiro lugar utilizando os parâmetros estimados com toda a amostra e a mesma probabilidade recorrendo aos parâmetros estimados utilizando um subconjunto de dados da amostra inicial.

Para todos os modelos o coeficiente de *Pearson* foi determinado e foram verificadas as seguintes hipóteses:

H_0 : As duas probabilidades não estão associadas.

H_1 : As duas probabilidades estão associadas (teste bilateral).

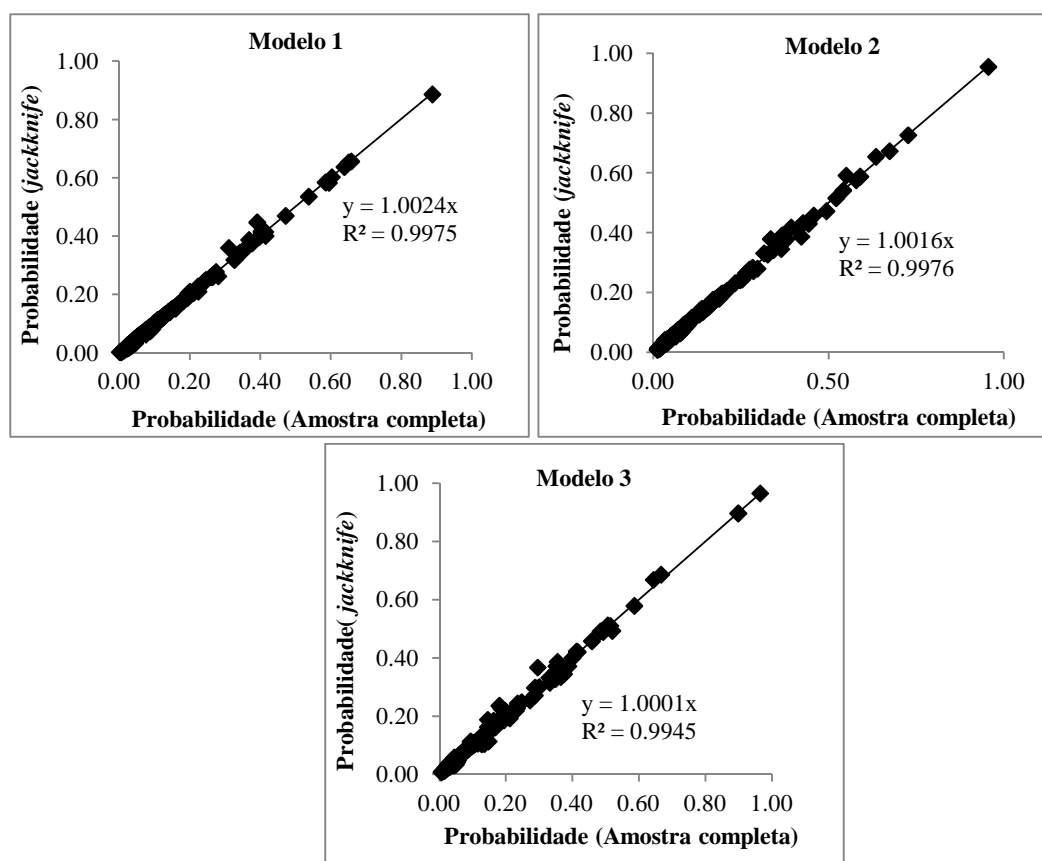


Figura 12. Correlações obtidas entre as probabilidades determinadas com a metodologia *jackknife* e as probabilidades iniciais.

Os resultados da Figura 12 mostram fortes correlações entre as duas probabilidades (coeficiente de *Pearson* superior a 0,9). O valor de prova obtido em qualquer um dos casos é inferior a 0,001 existindo evidência estatística de que as duas probabilidades estão associadas.

A conclusão final é que os modelos obtidos são robustos. O resultado obtido é tido como muito bom pois a amostra utilizada engloba investigadores de diferentes áreas científicas, concursos para dois tipos de posições na carreira académica e o peso atribuído ao desempenho científico na avaliação global varia de concurso para concurso.

O modelo com maior probabilidade de ser o melhor na explicação das decisões dos painéis de avaliação

O *ROLR* é ajustado a um conjunto de dados utilizando o critério da máxima verosimilhança. Para cada modelo a função da máxima verosimilhança pode ser utilizada no cálculo do Critério de Informação de *Akaike* (*AIC*) (Wagenmakers e Farrell 2004). Perante um conjunto de modelos aquele que possui o menor valor de *AIC* é considerado o melhor modelo.

O *AIC* é calculado utilizando a seguinte expressão:

$$AIC = 2k - 2 \log l(\beta) \quad (88)$$

onde k representa o número de parâmetros que definem o modelo e $\log l(\beta)$ é o valor da função de verosimilhança.

Os pesos de *Akaike* podem também ser determinados, já que fornecem informação sobre a probabilidade de cada modelo ser o melhor de entre o conjunto de modelos considerados. A sua forma de cálculo é a seguinte:

$$w_i = \frac{e^{(-0,5 \times \Delta_i)}}{\sum e^{(-0,5 \times \Delta_i)}} \quad (89)$$

Nesta equação o Δ_i representa a diferença entre o valor de *AIC* do modelo em causa e o *AIC* do modelo com o menor valor de *AIC*.

Os valores obtidos para o *AIC* e os pesos de *Akaike* são apresentados na Tabela 49.

Tabela 49. Valores obtidos para *AIC* e o w_i .

Modelo	<i>AIC</i>	w_i (%)
Modelo 2	253,06	65,8
Modelo 3	254,68	29,4
Modelo 1	258,30	4,8

Os resultados sugerem que o modelo 2 tem a maior probabilidade de ser o melhor. A probabilidade de este modelo ser o melhor na explicação das decisões tomadas pelos pares é cerca de duas vezes superior à obtida para o modelo 3 e cerca de 14 vezes superior à do modelo 1. As diferenças obtidas entre o modelo 2 e 1 eram já esperadas. Na Figura 11 é possível observar que o ρ do modelo 2 é ligeiramente superior ao do modelo 1. Na realidade o modelo 1 é o que apresenta sempre o menor valor para o ρ , excetuando o caso da variável B1. No entanto, os resultados que se mostram na Figura 11 não permitem uma escolha entre o modelo 2 e 3 uma vez que os valores de ρ são os mesmos para a variável dependente definida.

Note-se que o objetivo é obter um bom modelo utilizando o menor número possível de parâmetros. O *AIC* possui um termo de penalização que está relacionado com o número de parâmetros necessários para definir um modelo. Quanto maior o número de parâmetros maior o valor de *AIC*. O modelo 2 e 3 possuem valores para a função da máxima verosimilhança

muito semelhantes (-124,53 e -124,34, respetivamente), mas o número de parâmetros que definem o modelo 3 é maior resultando um valor de *AIC* superior para este modelo.

Influência dos indicadores bibliométricos na componente determinística

Com os modelos, agora, definidos é necessário estudar como é que um modelo responde perante pequenas/grandes alterações dos valores dos indicadores. Inicialmente foi estudado o comportamento esperado para os investigadores cujos indicadores tomam os valores de diversos percentis da amostra utilizada. A componente determinística de cada modelo é apresentada em baixo.

Modelo 1

$$V_{ni}=1,06PC_{ni} \text{ (90)}$$

Modelo 2

$$V_{ni}=0,40h_{nf_{ni}}+0,032HCD_{ni} \text{ (91)}$$

Modelo 3

$$V_{ni}=0,11NDF_{ni}+0,044HCD_{ni}+0,19NA_{mni} \text{ (92)}$$

Na Tabela 50 são apresentados os valores assumidos por estas componentes para os vários percentis considerados.

Tabela 50. Valores determinados para a componente determinística.

Modelo	Percentil						Média
	10	25	50	75	90	99	
Modelo 1	-1,67	-0,77	-0,03	0,74	1,44	2,62	0,16
Modelo 2	0,47	1,15	1,82	2,72	3,84	5,43	2,02
Modelo 3	1,07	2,05	2,97	4,12	5,55	8,71	3,23

No caso do modelo 1 o valor da componente determinística foi determinado considerando o valor que cada indicador que integra o *PC* toma no respetivo percentil para a amostra considerada.

Os dados apresentados na Tabela 50 permitem verificar, por exemplo, que perante um investigador A com uma componente determinística igual a 0,47 (percentil 10%) e um investigador B com uma componente determinística igual a 1,15 (percentil 25%) a probabilidade de A ser selecionado em primeiro lugar, se estes fossem os únicos candidatos, é de 33,6% (utilizando a expressão 76). A evolução da componente determinística é também mostrada na Figura 13 para cada um dos modelos.

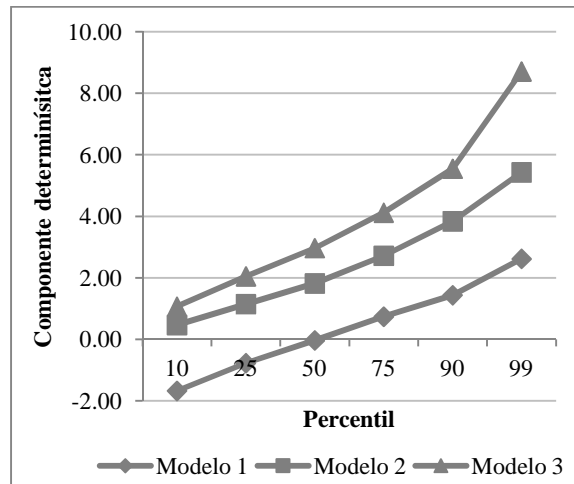


Figura 13. Evolução da componente determinística com o valor assumido em cada percentil pelos indicadores que definem os modelos.

Os dados apresentados anteriormente apenas permitem inferir sobre o resultado esperado para investigadores com determinadas características. No entanto a componente determinística é dependente de um conjunto de indicadores e portanto é essencial avaliar como é que um aumento/diminuição do valor do indicador influencia a componente. Isto é possível fazendo o uso de elasticidades. As elasticidades foram avaliadas para dois pontos: valores médios e medianos. Os resultados são apresentados nas Tabelas 51 e 52.

Tabela 51. Valores determinados para as elasticidades para o valor médio.

Modelo	Variável	$\partial y / \partial x()$	$eyex()(\%)$	$ey\partial x()(\%)$	$\partial yex()$
Modelo 1	<i>PC</i>	1,06	100	644	0,16
Modelo 2	<i>h_{nf}</i>	0,40	68	20	1,38
	<i>HCD</i>	0,032	32	1,6	0,64
Modelo 3	<i>NDF</i>	0,11	36	3,3	1,18
	<i>HCD</i>	0,044	27	1,4	0,88
	<i>NA_m</i>	0,19	36	5,8	1,17

Tabela 52. Valores determinados para as elasticidades para o valor mediano.

Modelo	Variável	$\partial y / \partial x()$	$eyex()(\%)$	$ey\partial x()(\%)$	$\partial yex()$
Modelo 1	<i>PC</i>	1,06	100	3311	-0,03
Modelo 2	<i>h_{nf}</i>	0,40	68	22	1,24
	<i>HCD</i>	0,032	32	2	0,58
Modelo 3	<i>NDF</i>	0,11	32	3,6	0,95
	<i>HCD</i>	0,044	27	1,5	0,80
	<i>NA_m</i>	0,19	41	6,3	1,22

Para esclarecer o significado dos valores apresentados nas Tabelas 51 e 52 em seguida explica-se recorrendo a um exemplo (Tabela 51, modelo 2 e variável h_{nf}) a leitura que deve ser efetuada dos valores obtidos:

$\partial y / \partial x = 0,40$ – o valor da componente determinística aumenta com o h_{nf} , este irá aumentar (para uma razão constante) 0,40 quando o h_{nf} aumenta uma unidade.

$eyex = 68\%$ – o valor da componente determinística aumenta com o h_{nf} , este irá aumentar 68% quando o valor de h_{nf} duplica (ou aumenta 100%, isto é, duas vezes o valor médio).

$ey\partial x = 20\%$ – o valor da componente determinística aumenta com o h_{nf} , este irá aumentar 20% quando o h_{nf} aumenta uma unidade.

$\partial yex = 1,38$ – o valor da componente determinística aumenta com o h_{nf} , este irá aumentar 1,38 quando o valor de h_{nf} duplica (ou aumenta 100%, isto é, duas vezes o valor médio).

Dos resultados apresentados nas Tabelas 51 e 52 o valor obtido para o $ey\partial x$ do modelo 1 é provavelmente aquele que causa mais curiosidade. Para este modelo, a componente determinística aumenta com o aumento de PC . O aumento é de 644%, para o caso em que se utiliza a média, e de 3311%, para a mediana, quando se incrementa uma unidade em PC . Especial atenção deve ser dada a esta variável, pois estamos perante uma variável que assume valores muito baixos para a média e mediana. Nestas condições, aumentar o seu valor em uma unidade tem um efeito considerável no incremento da componente determinística.

Os valores apresentados não mostram muitas diferenças quando o efeito de cada um dos indicadores sobre a componente determinística é estudado utilizando os valores médios ou medianos.

4.4 Conclusões

Ao longo deste capítulo analisou-se a viabilidade de se utilizarem indicadores bibliométricos para a obtenção de um modelo capaz de traduzir os resultados finais das avaliações por pares em concursos para Professor Associado e Catedrático nas universidades portuguesas.

A decisão fornecida por cada painel é a ordenação dos candidatos que resulta da avaliação de um conjunto de parâmetros relativos ao percurso profissional de cada candidato. As técnicas de regressão multivariada usuais não permitem tratar este tipo de variável resposta, facto que levou à exploração de regressões de escolha discreta.

Partindo de um conjunto inicial de doze indicadores e usando a metodologia estatística descrita foi de facto possível chegar a três modelos que descrevem as decisões finais dos painéis. Os modelos alcançados apresentam as seguintes características:

- O modelo 1 é definido apenas pela variável *PC* que é uma variável composta pela combinação linear de nove indicadores.
- O modelo 2 é definido pelos indicadores h_{nf} e *HCD*. Os dois indicadores estão relacionados com o impacto dos documentos publicados.
- O modelo 3 é definido por três indicadores, o *NDF*, *HCD* e NA_m . Estes indicadores estão relacionados com a quantidade, impacto e colaboração de cada candidato respetivamente.
- Dos três modelos definidos o modelo 2 é o que apresenta a maior probabilidade de ser o melhor (65,8%), isto é, o que melhor explicará as decisões finais dos painéis de avaliação.
- Verificou-se que os três modelos apresentam um comportamento semelhante relativamente à robustez.

Os modelos foram definidos utilizando um conjunto de concursos que diferem bastante entre si no que se refere ao desempenho científico de cada candidato, pois ocorreram em diferentes áreas. Note-se que não se conhece quais os critérios utilizados pelos painéis para avaliar o desempenho científico nestes concursos. Perante as características apresentadas pela amostra optou-se pela utilização de um conjunto de indicadores normalizados para descrever o desempenho científico e que o descrevem em várias dimensões. Face à ausência daquela informação, os resultados obtidos devem ser considerados satisfatórios.

O desempenho pedagógico e as atividades de transferência de tecnologia e gestão foram avaliados também, mas não foram considerados na definição do modelo, pois não foi possível obter informação que permitisse determinar indicadores que caracterizassem estes aspetos. Apesar de o desempenho científico representar o maior peso na avaliação final de cada candidato, para a maioria dos concursos existe um conjunto de variáveis não observáveis que não pode ser esquecido quando se avaliam e interpretam os resultados obtidos.

Os resultados obtidos apontam para a utilização da informação providenciada pelos indicadores bibliométricos como um instrumento auxiliar, já que fornecem informação útil e manejável, permitindo apoiar devidamente os avaliadores no processo de decisão.

Considera-se que a amostra utilizada não é de grandes dimensões, mas perante os resultados alcançados conclui-se que está “aberto o caminho” para estudos ao nível individual utilizando amostras maiores, outros indicadores, outras bases de dados e concursos onde o desempenho científico é o único parâmetro avaliado.

Levando em consideração o que foi mencionado anteriormente, considera-se que, como instrumento de apoio à decisão e no tipo de concursos aqui estudado, as regressões de escolha discreta podem ter um papel relevante no âmbito das técnicas bibliométricas.

Perante estas conclusões o passo seguinte será analisar o quão longe pode ir um modelo deste tipo como instrumento auxiliar da avaliação por pares. Este tema será abordado ao longo do próximo capítulo.

Capítulo 5

Poder de previsão do modelo

Por si só, a definição dos modelos não é suficiente para determinar a sua verdadeira utilidade como instrumento de apoio à decisão nas avaliações dos concursos. Para inferir sobre a aplicabilidade destes modelos é indispensável o estudo do seu poder de previsão. De facto, um modelo deste tipo não será certamente satisfatório se apresentar um baixo poder de previsão.

As decisões tomadas pelos avaliadores dos concursos aqui utilizados foram baseadas num conjunto de parâmetros que se acredita serem os mais importantes para descrever o desempenho profissional dos vários candidatos ao longo dos anos. Tal como foi previamente mencionado, na definição dos modelos apenas foram utilizados indicadores de desempenho científico. Dado que existe um conjunto de parâmetros que são avaliados pelo painel de cada concurso e que não foram considerados na construção dos modelos é importante estudar a sua aplicabilidade e o seu poder de previsão. Por um lado, um baixo poder de previsão poderá ser justificado pelo facto de não terem sido considerados estes parâmetros. Por outro, existem as limitações associadas à avaliação por pares e que foram mencionadas ao longo do Capítulo 1 (falta de concordância dentro do painel e o possível viés introduzido por algumas das características apresentadas pelos candidatos e pelo painel).

Assim neste capítulo procurou analisar-se o poder de previsão dos modelos. Note-se que nos estudos que têm vindo a ser efetuados ao longo dos últimos anos os modelos baseados em indicadores bibliométricos mostraram alguma correlação com as decisões dos painéis de avaliadores. No entanto, o poder de previsão dos modelos é pouco explorado e na maioria dos estudos o assunto não é abordado.

5.1 Amostra

O estudo do poder de previsão dos modelos foi efetuado utilizando os 171 candidatos dos 27 concursos. A ausência de um número considerável de concursos nos últimos anos levou a que o poder de previsão fosse estudado com os concursos utilizados na definição dos modelos.

5.2 Metodologia

Numa primeira fase, o poder de previsão foi estudado recorrendo a uma metodologia bastante simples. Foi determinada a percentagem de concursos para os quais o candidato colocado em primeiro lugar pelo júri possui uma probabilidade de ser escolhido em primeiro lugar superior à probabilidade esperada num cenário caracterizado pela aleatoriedade, isto é, um cenário em que a ordenação dos candidatos fosse obtida por sorteio. A informação disponibilizada pelos modelos permite ir mais longe e fazer o mesmo estudo para pares de candidatos, avaliados dois a dois. O conjunto de concursos disponíveis permitiu avaliar 426 pares.

Os pares foram contabilizados da forma como se ilustra na Figura 14.

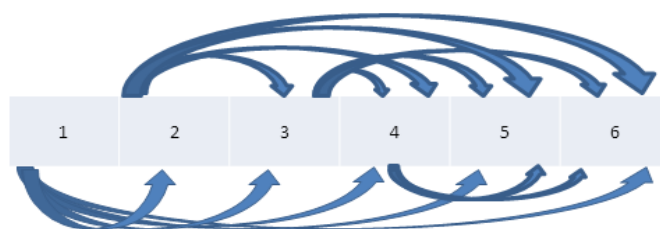


Figura 14. Contagem de pares para um concurso com seis candidatos e cuja ordenação é a conhecida.

Quando existem mais do que 4 candidatos a um concurso, os pares que representam os candidatos colocados na posição cinco ou inferior não foram considerados na avaliação do poder de previsão do modelo. Por exemplo o par 5-6 ou 5-7 não foram utilizados pelos motivos mencionados previamente no Capítulo 4. Isto justifica que na Figura 14 apenas se apresentem os pares formados pelos candidatos 1 até 4 relativamente a todos os outros candidatos (14 pares).

Com este procedimento conseguiu-se uma primeira visão acerca do quão longe podem ir estes modelos como instrumento de apoio à decisão na avaliação por pares.

Função de distribuição de probabilidade associadas aos resultados dos modelos

Tal como foi referido no ponto anterior, o poder de previsão dos modelos foi inicialmente avaliado utilizando um processo simplista. Informação mais segura pode ser obtida a partir da função de distribuição de probabilidade associada a cada modelo e recorrendo ao número de pares formados pelos candidatos. Na função distribuição de probabilidade será estudada a variável Y que representa o número de pares cuja ordenação relativa é prevista corretamente pelo modelo. Com estas distribuições pretende estimar-se o valor esperado de acertos para cada um dos modelos.

A metodologia utilizada na construção dos modelos permite determinar probabilidades para subconjuntos de alternativas. Assim para cada par é possível determinar a probabilidade de cada candidato ser selecionado em primeiro lugar. Um exemplo deste cálculo é apresentado em seguida:

Exemplo

Imagine-se um concurso com três candidatos (C1, C2, C3) em que a probabilidade de cada um ficar em primeiro lugar segundo um dos modelos é de 53,8%, 20,2% e 26%, respetivamente. O candidato C1 foi colocado em primeiro lugar pelo painel e os candidatos C2 e C3 em segundo e terceiro lugar respetivamente. Para o par formado por C1 e C2 a probabilidade de C1 ficar em primeiro lugar é de 72,7% e a de C2 é de 27,3%.

Para cada par existem dois resultados possíveis: um deles será denominado de “sucesso” com probabilidade de acontecer igual a p e o outro de “não sucesso” com probabilidade de acontecer igual a $1-p$. Aqui “sucesso” significa que a previsão fornecida pelo modelo coincide com o resultado da avaliação por pares. A probabilidade de “sucesso” corresponde à probabilidade de o candidato colocado na melhor posição no par pelo painel ser selecionado

em primeiro lugar segundo as previsões do modelo. A probabilidade de “sucesso” difere de par para par e, como tal, não foi possível aplicar nenhuma das distribuições de probabilidades discretas conhecidas. A situação é semelhante à da distribuição Binomial, com a diferença de que a probabilidade associada a cada tentativa não ser sempre a mesma. Perante esta situação a solução encontrada passou pela utilização de um processo de simulação. No processo de simulação foi utilizado o método de Monte Carlo.

O simulador gera um conjunto de números aleatórios a partir dos quais se estudam as variáveis de interesse. Em cada simulação é necessário indicar o número de iterações que devem ser realizadas. Normalmente, um número elevado de iterações é requerido por simulação para garantir a estabilidade do processo. Para 100 000 iterações verificou-se que a estabilidade era alcançada.

As distribuições obtidas a partir da simulação apresentavam uma forma muito próxima da distribuição Normal. Nesta situação e dado que o número de observações é muito elevado, foi possível utilizar a distribuição Normal para inferir sobre os resultados obtidos. Assim, a distribuição Normal foi ajustada aos dados, recorrendo-se ao teste de *Kolmogorov-Smirnov* para verificar a razoabilidade desta aproximação. A simetria e o achatamento foram também analisados para cada uma das distribuições.

Comparação das previsões dos modelos com as produzidas num cenário caracterizado pela aleatoriedade

A comparação das previsões fornecidas pelos modelos com as previsões de um cenário caracterizado pela aleatoriedade (onde nada seria dito sobre os candidatos e os mesmos seriam seleccionados por sorteio) permite verificar se os modelos definidos acrescentam, de facto, informação.

Para o cenário aleatório assumiu-se que o número de pares corretos (X) (ou seja nos quais este cenário acerta) segue uma distribuição Binomial. Numa situação em que o número de observações, neste caso o número de pares n , é muito elevado a distribuição Binomial pode ser aproximada pela distribuição Normal (Box, Hunter e Hunter 2005).

Se o número de observações é superior a 5 a aproximação utilizando a distribuição Normal é adequada quando (Box, Hunter e Hunter 2005):

$$\left(\frac{1}{\sqrt{n}}\right) \times \left(\left(\frac{\sqrt{1-p}}{p} \right) - \left(\sqrt{\frac{p}{1-p}} \right) \right) < 0,3 \quad (93)$$

No caso em apreço n é igual a 426 e p é a probabilidade de o candidato colocado em primeiro lugar pelo painel ser escolhido num cenário caracterizado pela aleatoriedade. Dado que se estão a considerar pares a cada candidato está associada uma probabilidade de 0,5.

Se a inequação representada pela expressão 93 é verdadeira, então a variável X segue uma distribuição caracterizada pelos seguintes parâmetros:

$$X \sim N(n \times p; \sqrt{n \times p(1-p)}) \quad (94)$$

Escrevendo de uma outra forma o número de pares previstos corretamente (X) segue uma distribuição com média igual a $n/2$ e desvio padrão igual a $\sqrt{n}/2$.

$$X \sim N\left(\frac{n}{2}; \frac{\sqrt{n}}{2}\right) \quad (95)$$

Para cada modelo obteve-se também que o número de pares previstos corretamente (Y) segue uma distribuição Normal (como se verá posteriormente):

$$Y \sim N(\mu; \sigma) \quad (96)$$

Finalmente, verificou-se a probabilidade de as previsões fornecidas num cenário aleatório serem superiores às fornecidas pelos modelos definidos. Para cada modelo foi então calculada a seguinte probabilidade:

$$P(X > Y) = P(X - Y > 0) \quad (97)$$

Poder de previsão quando apenas a produção científica é avaliada pelos pares

Os modelos foram definidos utilizando apenas as variáveis que descrevem o desempenho científico dos candidatos. Na realidade, o painel avalia outros aspetos para além do desempenho científico. O peso do desempenho científico do candidato representa entre 60% e 70% do total da avaliação final em 74% dos concursos em análise. Para os restantes (28%), o peso atribuído ao desempenho científico varia entre 45% e 55%. Medir o desempenho dos candidatos que não no aspeto científico é uma tarefa complexa, uma vez que a obtenção da informação nem sempre é viável e no caso em estudo isso não foi mesmo possível. O facto de estes aspetos não terem sido considerados na definição do modelo tem certamente consequências negativas. A influência do termo aleatório será maior e a precisão associada aos resultados obtidos será menor. Descreve-se em seguida o procedimento adotado para o demonstrar.

Na expressão abaixo, que descreve a utilidade associada a cada alternativa, introduziu-se um novo termo relacionado com os restantes aspetos avaliados pelo painel e que não podem ser descritos por indicadores bibliométricos ($\delta_{nj2}y_{nj2}$):

$$U_{nj} = a_{nj1}\beta_{nj1}x_{nj1} + b_{nj2}\delta_{nj2}y_{nj2} + \varepsilon_{nj} \quad (98)$$

Nesta expressão, o termo $\beta_{nj1}x_{nj1}$ representa a dimensão desempenho científico que foi avaliada utilizando indicadores bibliométricos, a_{nj1} o peso atribuído em cada concurso ao desempenho científico, b_{nj2} o peso atribuído aos restantes aspetos avaliados pelo júri e ε_{nj} o termo estocástico definido anteriormente.

Com este novo termo na expressão da utilidade a probabilidade de o decisor escolher a alternativa i passa a ser representada por:

$$P_{ni}=P((V_{ni1}+V_{ni2})-(V_{nj1}+V_{nj2})>\varepsilon_{nj}-\varepsilon_{ni}, \forall j \neq i) \quad (99)$$

onde se assumiu V_{nj1} para representar $\beta_{nj1}x_{nj1}$ e V_{nj2} para representar $\delta_{nj2}y_{nj2}$.

Expressando as probabilidades sobre a densidade das diferenças entre os termos do erro:

$$P_{ni} = \int I(\varepsilon_{nj}-\varepsilon_{ni} < (V_{ni1}+V_{ni2})-(V_{nj1}+V_{nj2}) \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (100)$$

Considerando que apenas a diferença entre as utilidades é relevante, que a normalização ocorre automaticamente e que os ε_{nj} são independentes, a função distribuição conjunta de todos os $j \neq i$ resulta do produto das funções de distribuição individuais:

$$P_{ni|\varepsilon_{ni}} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + (V_{ni1}+V_{ni2})-(V_{nj1}+V_{nj2}))}} \quad (101)$$

Uma vez que ε_{ni} e V_{n2} não são conhecidos, é possível calcular aquela probabilidade utilizando a seguinte expressão e introduzindo o peso que cada um dos parâmetros avaliados possui na decisão final (a_{nj1} e b_{nj2}):

$$P_{ni|\varepsilon_{ni}} = \int f(V_{ni2}) dV_{ni2} \int f(V_{nj2}) dV_{nj2} \int \prod_{j \neq i} e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} e^{-e^{-(\varepsilon_{ni} + a_{n1}(V_{ni1}-V_{nj1}) + b_{n2}(V_{ni2}-V_{nj2}))}} d\varepsilon_{ni} \quad (102)$$

$$P_{ni|\varepsilon_{ni}} = \int f(V_{i2}) dV_{i2} \int f(V_{j2}) dV_{j2} \frac{e^{a_{n1}V_{ni1} + b_{n2}V_{ni2}}}{e^{a_{n1}V_{ni1} + b_{n2}V_{ni2}} + e^{a_{n1}V_{nj1} + b_{n2}V_{nj2}}} \quad (103)$$

$$P_{ni|\varepsilon_{ni}} = \int f(V_{i2}) dV_{i2} \int f(V_{j2}) dV_{j2} \frac{1}{1 + e^{a_{n1}(V_{i1}-V_{j1}) + b_{n2}(V_{i2}-V_{j2})}} \quad (104)$$

Se o coeficiente b_{n2} assumir o valor zero, isto é, se o desempenho científico for o único aspeto avaliado, o cálculo da expressão é direto:

$$P_{ni} = \frac{1}{1 + e^{a_{n1}(V_{ni1}-V_{nj1})}} \quad (105)$$

Quando o valor de b_{n2} é diferente de zero o cálculo da expressão 104 é complexo porque não é conhecida informação acerca de V_2 . Perante isto algumas hipóteses de trabalho foram elaboradas.

Aqui considerou-se que o termo estocástico engloba principalmente os atributos não observados pelo analista que estão relacionados de alguma forma com V_1 e V_2 , isto é, existem atributos relacionados com os parâmetros V_1 e V_2 que contribuíram para a formulação das decisões de cada painel, mas que o analista não conseguiu identificar. Perante isto

considerou-se que a dispersão de ε_n está relacionada com a dispersão em V_1 e V_2 . Conhecendo as distribuições de V_1 e V_2 é possível inferir sobre a dispersão de ε_n .

As hipóteses de trabalho consideradas estão relacionadas com as distribuições de V_1 e V_2 .

Hipótese 1: V_1 e V_2 possuem distribuições Normais

$$V_1 \sim N(\mu_{V_1}; \sigma_{V_1}) \quad (106)$$

$$V_2 \sim N(\mu_{V_2}; \sigma_{V_2}) \quad (107)$$

É também necessário dispor de informação sobre a dispersão destas distribuições. Considerando que os painéis de avaliadores colocam o mesmo esforço na avaliação V_1 e V_2 e que não existem motivos que fundamentem um comportamento oposto, admitiu-se que os dois termos (V_1 e V_2) possuem a mesma dispersão.

Hipótese 2: V_1 e V_2 têm a mesma dispersão

$$\sigma_{V_1} = \sigma_{V_2} = \sigma_V \quad (108)$$

Se a dispersão de ε_n resultar da dispersão de V_1 e de V_2 então:

$$\sigma_{\varepsilon_n} = \sqrt{\left((a_{n1}\sigma_{V_1})^2 + (b_{n2}\sigma_{V_2})^2\right)} \quad (109)$$

$$\sigma_{\varepsilon_n} = \sigma_V \sqrt{((a_{n1})^2 + (b_{n2})^2)} \quad (110)$$

A distribuição Normal foi utilizada para inferir acerca da dispersão do termo estocástico, quando, de facto, é considerado que este termo possui uma distribuição *idd* de valor extremo. Observando a expressão 99 conclui-se que no cálculo da probabilidade apenas a diferença de utilidades é importante. Perante duas alternativas j e k a diferença entre os termos estocásticos origina um erro, ε_n , que possui uma distribuição logística. A distribuição logística tem uma forma bastante semelhante à distribuição Normal mas possui caudas mais “pesadas” (maior curtose) permitindo controlar melhor as observações mais atípicas nas amostras. Perante a semelhança das distribuições considerou-se que a hipótese de que as variáveis V_1 e V_2 seguem aproximadamente uma distribuição Normal é aceitável.

Com base nas hipóteses de trabalho consideradas foi possível determinar o intervalo de valores que o σ_{ε_n} poderá assumir.

Se $b_n=1$:

$$\sigma_{\varepsilon_n} = \sigma_V \quad (111)$$

Se $b_n = \frac{1}{2}$

$$\sigma_{\varepsilon_n} = \sigma_V \frac{1}{\sqrt{2}} \quad (112)$$

Adicionalmente $\sigma_{\varepsilon_n} \in \left[\frac{1}{\sqrt{2}}; 1 \right] \times \sigma_V$

No modelo *ROLR* a utilidade é normalizada utilizando a variância do termo estocástico como se referiu anteriormente na secção destinada à metodologia no Capítulo 4. Assim os coeficientes obtidos refletem o impacto das variáveis observáveis relativamente ao desvio padrão das variáveis não observáveis (Train 2009). As variâncias dos erros no tipo de modelo aqui adotado são normalizadas para $\frac{\pi^2}{6}$, $\text{var}(\varepsilon_{nj}) = 1,64$ (ver descrição deste aspeto na secção destinada à metodologia do Capítulo 4). Considerando isto a expressão da utilidade possuirá a seguinte forma:

$$U_{nj} = a_{nj} \frac{\beta_{nj1}^*}{\sigma_{\varepsilon_{nj}}} \frac{\pi}{\sqrt{6}} x_{nj1} + b_{nj} \frac{\delta_{nj2}^*}{\sigma_{\varepsilon_{nj}}} \frac{\pi}{\sqrt{6}} y_{nj2} + \varepsilon_{nj} \quad (113)$$

Se

$$\sigma_{\varepsilon_n} \in \left[\frac{1}{\sqrt{2}}; 1 \right] \times \sigma_V \quad (114)$$

Então para o intervalo de variação de σ_{ε_n} :

$$\frac{\pi}{\sqrt{6}} = \frac{1}{\sqrt{2}} \sigma_{V_2} \quad (115)$$

$$\frac{\pi}{\sqrt{6}} = 1 \sigma_{V_2} \quad (116)$$

Daqui resulta que:

$$\sigma_{V_2} \in \left[\frac{\pi}{\sqrt{6}}; \frac{\pi}{\sqrt{3}} \right] \quad (117)$$

Com informação acerca dos valores que o σ_{V_2} poderá assumir é agora possível afirmar que :

$$V_2 \sim N\left(0; \frac{\pi}{\sqrt{6}}\right) \quad (118)$$

O menor valor do intervalo foi selecionado pois pretende-se mostrar o melhor resultado que pode ser obtido com a utilização dos modelos. À medida que o desvio padrão diminui, a influência do termo estocástico também diminui e os resultados serão mais satisfatórios.

O cálculo da expressão 104 foi efetuado utilizando o método de Monte Carlo. Os seguintes passos foram necessários:

- 1) Considerando na expressão 104 que:

$$K = e^{a_{n1}(V_{ni1}-V_{nj1})} \quad (119)$$

Se b_{n2} é igual a zero na expressão 104 aplica-se então a expressão 105:

$$e^{(V_{ni1}-V_{nj1})} = \left(\frac{1}{P_{ni}} - 1 \right) \quad (120)$$

Assim é possível obter:

$$K = \left(\frac{1}{P_{ni}} - 1 \right)^{a_{n1}} \quad (121)$$

- 2) Fixaram-se valores de P_{ni} (probabilidade de o candidato ser selecionado em primeiro lugar se apenas o desempenho científico fosse avaliado) e determinou-se o valor de K .
- 3) Conhecendo os valores de K , a média e desvio padrão de V_2 o valor do integral na expressão 104 foi determinado utilizando o método de Monte Carlo.

Com as novas probabilidades determinadas foi possível estimar novas funções distribuição de probabilidades e, a partir delas, calcular o valor esperado do número de pares previstos corretamente quando apenas o desempenho científico é avaliado.

O rigor da estimativa de σ_{V_2}

Com os dados empíricos disponíveis foi também possível determinar o rigor da estimativa de σ_{V_2} definida anteriormente. Para isso admite-se a hipótese de que V_1 e V_2 possuem a mesma dispersão:

$$\sigma_{V_1} = \sigma_{V_2} \quad (122)$$

Com a inserção do novo termo, relacionado com os restantes parâmetros avaliados, a função utilidade passa a ser descrita pela expressão 98. Imaginando que o termo V_1 é composto por duas variáveis (x_1 e x_2) o efeito deste termo sobre a dispersão da utilidade total é representado pela expressão descrita em seguida:

$$a_{nj} \left(\beta_{nj1} \sigma_{x_{nj1}} + \beta_{nj2} \sigma_{x_{nj2}} \right) \quad (123)$$

O efeito do termo V_2 na dispersão da utilidade total é representado pela seguinte expressão:

$$b_{nj}\delta_{nj2}\sigma_{y_{nj2}} \quad (124)$$

Os valores de β e δ são normalizados considerando a variância do termo estocástico (Train 2009):

$$\beta_{nj1} = \frac{\beta_{nj1}^*}{\sigma_{\varepsilon nj}} \times \frac{\pi}{\sqrt{6}} \quad (125)$$

$$\beta_{nj2} = \frac{\beta_{nj2}^*}{\sigma_{\varepsilon nj}} \times \frac{\pi}{\sqrt{6}} \quad (126)$$

$$\delta_{nj2} = \frac{\delta_{nj2}^*}{\sigma_{\varepsilon nj}} \times \frac{\pi}{\sqrt{6}} \quad (127)$$

Se $\sigma_{V_1} = \sigma_{V_2}$ o desvio padrão associado ao termo V_2 pode ser determinado utilizando a seguinte expressão:

$$\left(\beta_{nj1}\sigma_{x_{nj1}} + \beta_{nj2}\sigma_{x_{nj2}} \right) = \sigma_{V_2} \quad (128)$$

A determinação do valor de σ_{V_2} para cada modelo permite analisar a diferença entre os valores médios esperados para o número de pares previstos corretamente a partir da simulação e os valores que seriam obtidos utilizando os modelos.

5.3 O sucesso dos modelos na previsão dos resultados

Train 2009 (Capítulo 3) menciona que, em muitas situações e de modo errado, na avaliação dos resultados produzidos pelos modelos de escolha discreta é determinada a percentagem de casos previstos corretamente. No cálculo desta estatística é considerada a alternativa que possui maior probabilidade de ser escolhida em primeiro lugar e posteriormente é verificado se a mesma foi escolhida pelo decisor como a melhor. Neste cálculo considera-se que a escolha do decisor é baseada na alternativa para a qual o modelo determina a maior probabilidade quando isto não corresponde à realidade. Na verdade, o analista não possui informação suficiente para determinar a seleção de cada decisor, mas apenas informação suficiente que lhe permite determinar a probabilidade associada a cada alternativa. No caso em estudo, com as probabilidades obtidas está-se a considerar que se cada concurso fosse repetido inúmeras vezes cada alternativa (candidato) seria escolhido y vezes.

Neste trabalho adotou-se um procedimento diferente. Na Tabela 53 apresenta-se a percentagem de casos em que a probabilidade de o candidato colocado na melhor posição pelo painel ser selecionado em primeiro lugar é superior aquela que seria esperada num cenário onde nada se sabe acerca de cada candidato e a seleção é feita aleatoriamente.

Tabela 53. Comparação dos modelos com um cenário aleatório.

Modelo	$P_{RI}(\%)$	$P_{Pares}(\%)$
1	78	75
2	70	75
3	63	76

Na Tabela 53 P_{RI} é a percentagem de concursos em que o candidato colocado na posição um pelo painel de avaliação possui uma probabilidade de ser selecionado em primeiro lugar superior à probabilidade que teria se a seriação fosse elaborada à sorte. Na segunda coluna da Tabela 53, P_{Pares} indica a percentagem de pares (do total de 426) em que o candidato colocado na melhor posição no par pelo painel possui uma probabilidade de ser escolhido em primeiro lugar superior à probabilidade de um sorteio feito aleatoriamente (probabilidade igual a 50%). Considera-se que os valores daquelas percentagens são elevados para todos os modelos nas duas situações estudadas, uma vez que na parametrização do modelo apenas se consideraram variáveis relacionadas com um dos parâmetros avaliados (vertente desempenho científico).

Na Figura 15 representam-se os histogramas das frequências das probabilidades (P_{ni}) associadas aos candidatos colocados na melhor posição no respetivo par pelo painel de avaliação.

A observação da Figura 15 mostra de facto uma maior concentração dos pares do lado direito do histograma, isto é, para probabilidades superiores a 50% o que evidencia a superioridade dos modelos em relação a um cenário onde os candidatos são ordenados aleatoriamente.

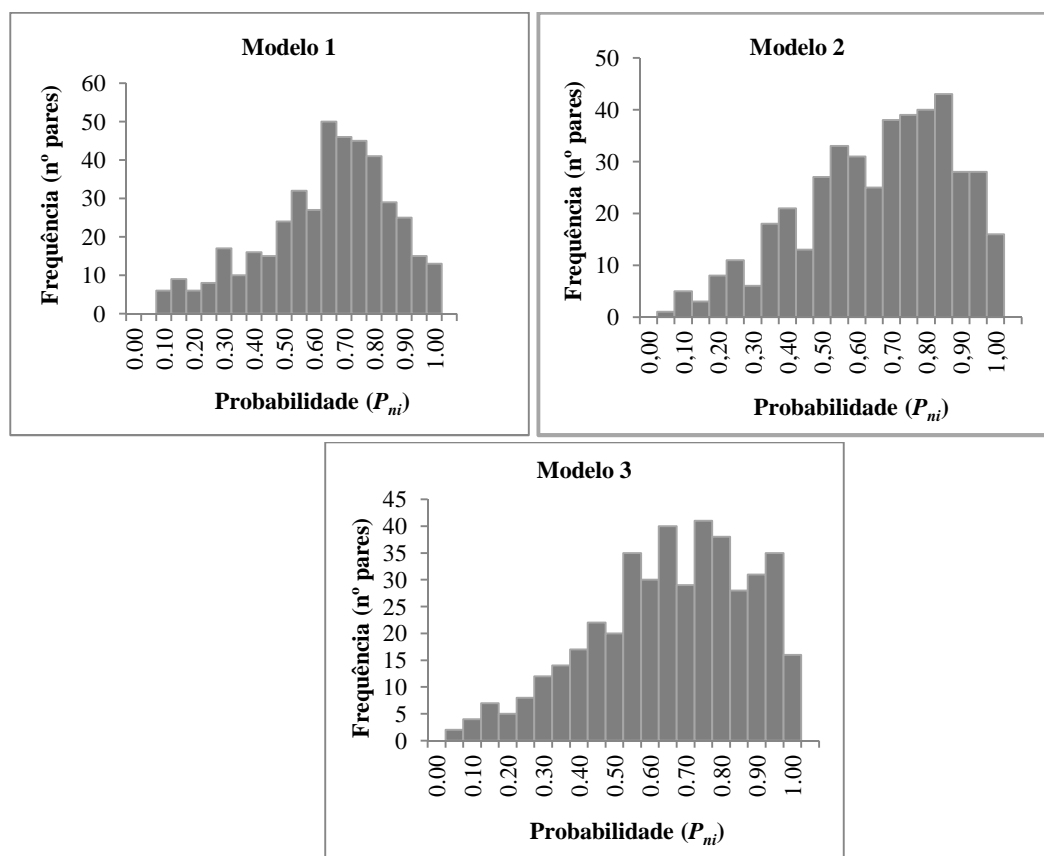


Figura 15. Histograma de distribuição de probabilidades de P_{ni} .

5.4 Função de distribuição de probabilidades

A avaliação da função distribuição de probabilidades associada a cada modelo permite ao analista inferir sobre o rigor das respectivas previsões. As distribuições obtidas e os parâmetros que as caracterizam são apresentados na Figura 16. Com estas distribuições pretende determinar-se qual o número médio de pares previstos corretamente que podem ser esperados com a utilização destes modelos.

As distribuições apresentadas na Figura 16 apresentam uma forma muito próxima da distribuição Normal. No caso aqui analisado as distribuições representam a probabilidade de ocorrer um determinado resultado para a variável discreta “número de pares previstos corretamente”. Numa distribuição Binomial a probabilidade de sucesso associada a cada evento é sempre a mesma para todas as tentativas. Tal como já se referiu, em cada tentativa a probabilidade de sucesso é diferente e portanto não poderá ser considerado que a distribuição Binomial descreve corretamente o comportamento dos dados. No entanto, dado que o número de observações é elevado, a distribuição Normal foi ajustada aos dados de modo a ser possível inferir sobre os resultados obtidos por simulação.

O teste de *Kolmogorov-Smirnov* foi utilizado para verificar as seguintes hipóteses:

H_0 : A amostra provém de uma distribuição Normal.

H_1 : A amostra não provém de uma distribuição Normal.

A aplicação do teste a cada um dos modelos levou à rejeição da hipótese nula, isto é, não existe evidência estatística de que os dados provêm de uma distribuição Normal ($p < 0,001$). No entanto, este teste apresenta uma elevada sensibilidade (potência) para amostras de grandes dimensões, tendo capacidade para detetar desvios mínimos em relação à distribuição assumida (neste caso distribuição Normal), rejeitando assim a hipótese. Uma alternativa passa pela análise do histograma de frequência dos dados acompanhando-o de informação acerca do achatamento e simetria. Na Figura 17 estão representados os histogramas com a curva Normal sobreposta para cada um dos modelos.

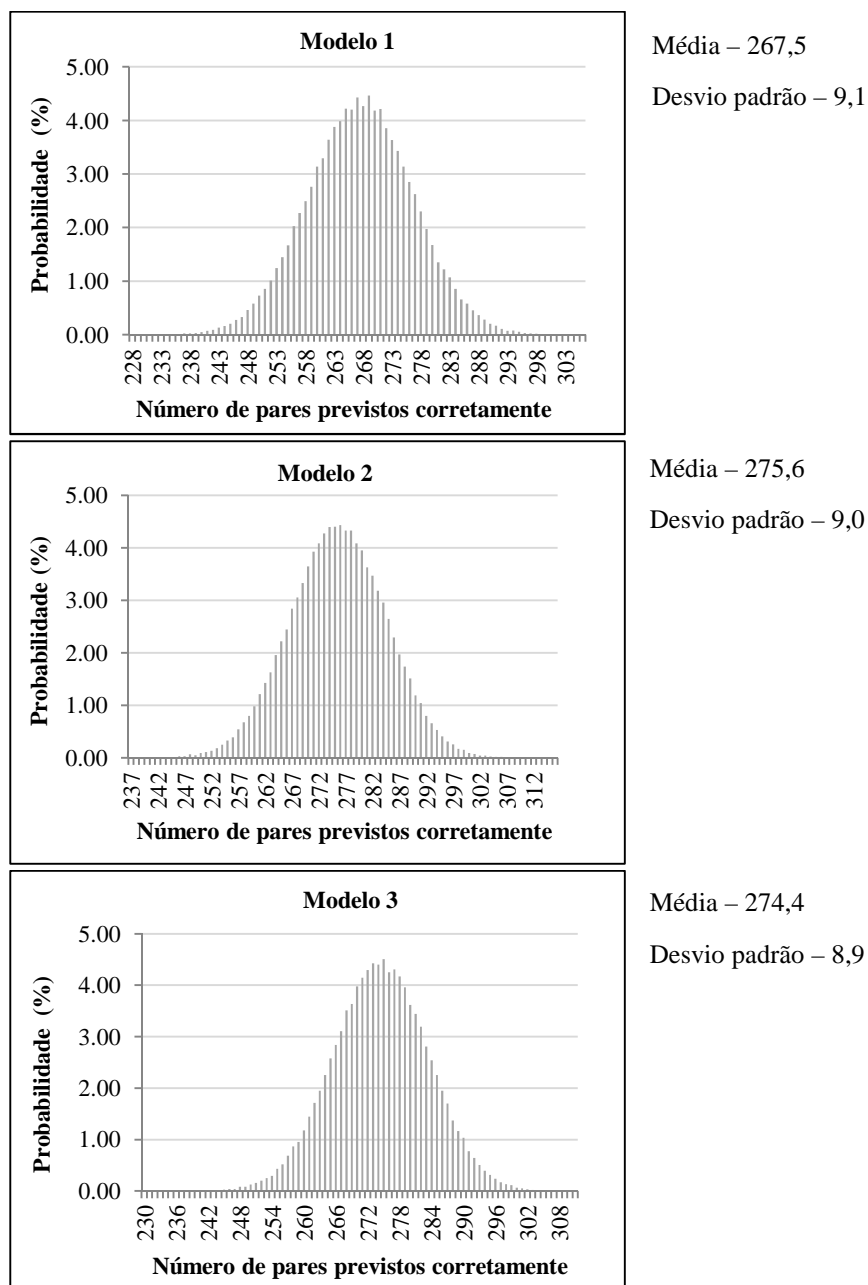


Figura 16. Função distribuição de probabilidades obtidas para cada modelo.

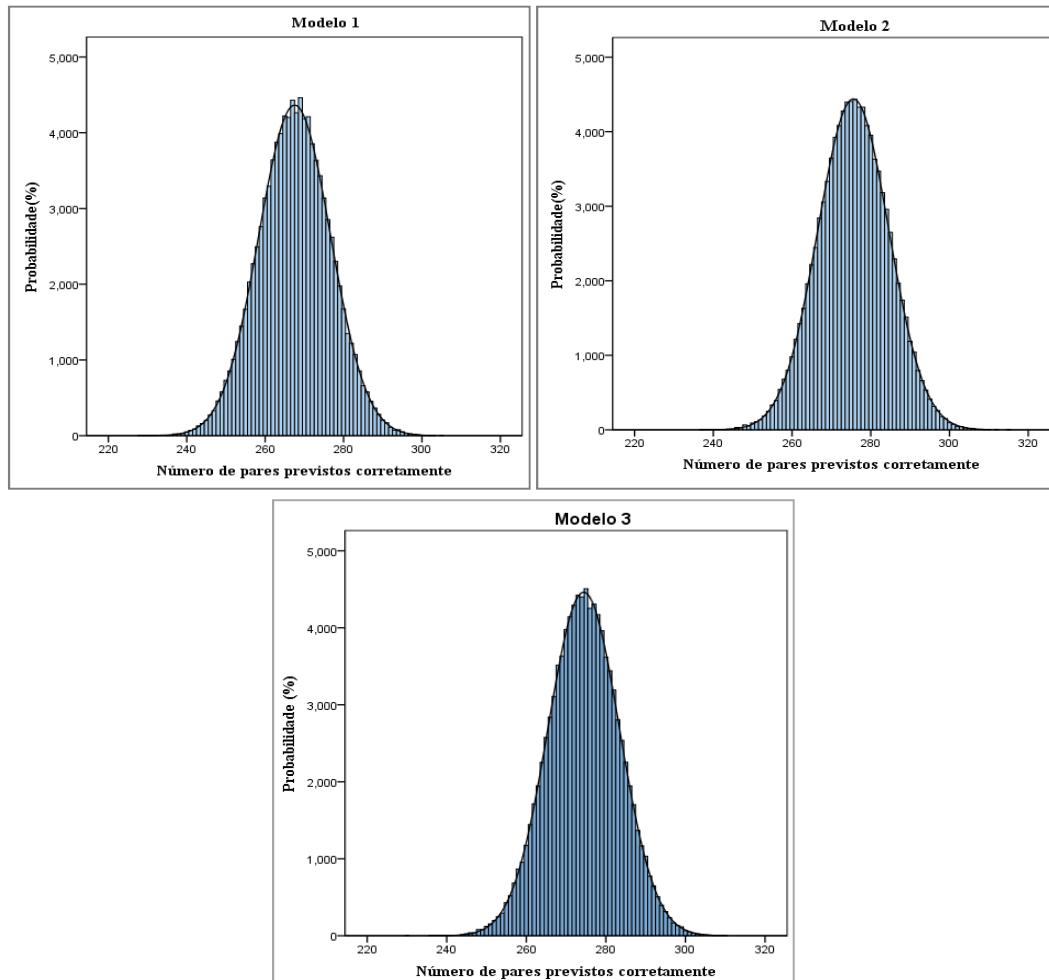


Figura 17. Ajuste da distribuição Normal a cada um dos modelos.

Observando a Figura 17 é possível constatar que, de facto, os dados provêm de uma distribuição muito próxima da Normal. Os valores obtidos para a simetria e a curtose (achatamento) de cada uma das distribuições, na Tabela 54, reforçam estas conclusões.

Tabela 54. Valores de simetria, curtose, média e desvio padrão obtidos para as distribuições de cada modelo.

Modelo	Simetria	Curtose	Média	Desvio padrão
Modelo 1	-0,014	0,012	267,5	9,1
Modelo 2	-0,028	0,003	275,6	9,0
Modelo 3	-0,028	0,000	274,4	8,9

Em todos os modelos os valores da simetria são próximos e indicam que as distribuições são assimétricas negativas, isto é, possuem a cauda à esquerda mais alongada. Os valores são, no entanto, muito próximos de zero e portanto as caudas obtidas são pequenas.

Os valores de curtose são também muito baixos. Para o modelo 1 e 2 os valores determinados sugerem que as distribuições são leptocúrticas (mais afunilada e concentrada que a distribuição Normal padrão) enquanto para o modelo 3 o curtose assume o valor zero e, portanto, a distribuição é mesocúrtica isto é possui o mesmo achatamento que a distribuição Normal.

Os resultados mostram que o valor esperado para o número de pares previstos corretamente é de $63\% \pm 2\%$, $65\% \pm 2\%$ e $64\% \pm 2\%$ para o modelo 1, 2 e 3, respetivamente. Quando comparado com o esperado num cenário onde não existe disponível informação acerca dos candidatos este resultado pode ser considerado bom (para um cenário como o descrito é esperado 50% de acerto).

5.5 Comparação das previsões dos modelos com um cenário aleatório

Numa situação onde não exista informação disponível acerca do desempenho científico dos candidatos todos têm a mesma probabilidade de serem escolhidos. Os modelos definidos apenas serão úteis se conseguirem fornecer mais informação do que aquela que está disponível num cenário caracterizado pela aleatoriedade.

Uma vez que se verificou a condição expressa pela inequação 93 apresentada na secção da metodologia, para o cenário caracterizado pela ausência de informação acerca dos candidatos considerou-se que o número de pares previstos corretamente (X) segue uma distribuição Normal com os seguintes parâmetros:

$$X \sim N\left(\frac{426}{2}; \frac{\sqrt{426}}{2}\right) \quad (129)$$

Como se verificou anteriormente, o número de pares previstos corretamente pelos modelos também segue uma distribuição Normal:

Modelo 1

$$Y_{Modelo1} \sim N(267,5; 9,1) \quad (130)$$

Modelo 2

$$Y_{Modelo2} \sim N(275,6; 9,0) \quad (131)$$

Modelo 3

$$Y_{Modelo3} \sim N(274,4; 8,9) \quad (132)$$

onde Y representa o número de pares previstos corretamente por cada modelo.

Conhecendo os parâmetros que caracterizam a função distribuição de probabilidade em ambas as situações é possível determinar a probabilidade de X ser superior a Y .

$$P(X > Y) = P(X - Y > 0) \quad (133)$$

A subtração de dois conjuntos de dados com distribuição Normal origina um novo conjunto de dados que segue também uma distribuição Normal, neste caso com os seguintes parâmetros:

$$(X-Y) \sim N\left(\frac{n}{2} - \mu; \sqrt{\left(\left(\frac{\sqrt{n}}{2}\right)^2 + (\sigma)^2\right)}\right) \quad (134)$$

Na Tabela 55 estão representados os vários parâmetros das funções distribuição de probabilidade.

Para cada modelo verificou-se que a probabilidade de as previsões disponíveis numa situação em que nada se sabe acerca do desempenho científico dos candidatos serem superiores às previsões fornecidas pelos modelos é efetivamente muito baixa (inferior a 0,01%). Este resultado permite concluir que os modelos obtidos conseguem fornecer mais informação do que aquela que está disponível numa situação em que os candidatos são selecionados aleatoriamente. Portanto, a utilização destes modelos como um instrumento de apoio será uma mais valia no processo de decisão.

Tabela 55. Média e desvio padrão para cada função distribuição de probabilidade.

Variável	Modelo	Média	Desvio padrão
Y	Modelo 1	267,5	9,1
	Modelo 2	275,6	9,0
	Modelo 3	274,4	8,9
X	Seleção aleatória	213	10,3
(X-Y)	Modelo 1	-54,5	13,5
	Modelo 2	-62,6	13,5
	Modelo 3	-61,4	13,4

5.6 Poder de previsão dos modelos quando apenas a produção científica é avaliada

Nesta secção é mostrada a influência dos restantes parâmetros avaliados pelos painéis dos concursos no poder de previsão de cada um dos modelos. Para o caso em que apenas o desempenho científico é avaliado espera-se que a informação fornecida pelos modelos seja melhor. De facto, a influência do termo estocástico, que não é observado pelo analista, afetará menos o poder de previsão de cada modelo. No edital de cada concurso é indicado o peso atribuído ao desempenho científico na avaliação final de cada candidato. Esta informação foi utilizada para derivar uma nova função para o cálculo da probabilidade de um dado candidato ser selecionado em primeiro lugar. A integração da função derivada foi efetuada utilizando o método de Monte Carlo já que, como se pode ver pela expressão 104 apresentada na secção destinada à metodologia, não é possível resolver a mesma analiticamente.

Na Figura 18 representa-se a relação entre a probabilidade de um candidato ser selecionado em primeiro lugar de acordo com o peso atribuído ao desempenho científico na avaliação final (avaliação mista) e a probabilidade esperada se apenas fosse avaliado o

desempenho científico (eixo das ordenadas). As curvas na figura representam a correlação consoante o peso atribuído na avaliação final aos parâmetros avaliados pelo painel e que não foram considerados na definição dos modelos.

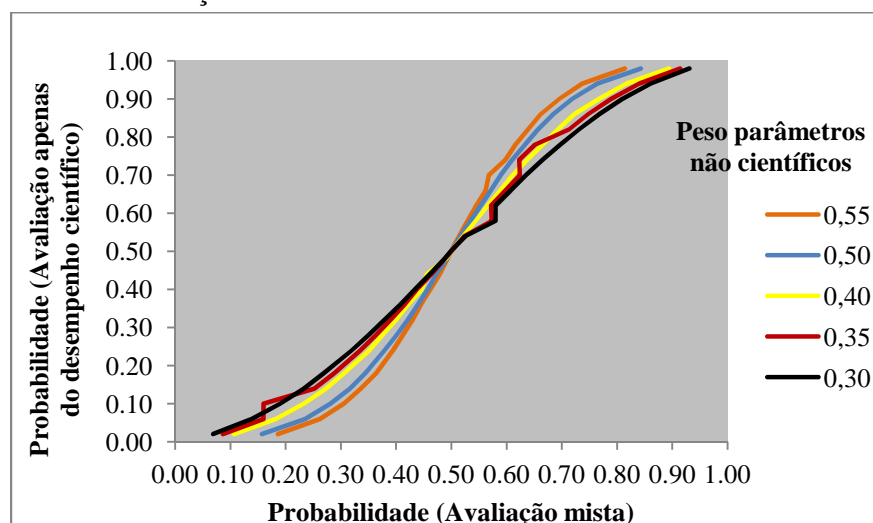


Figura 18. Relação entre a probabilidade de um candidato ser selecionado em primeiro se apenas o desempenho científico é avaliado e a probabilidade para o caso real onde a avaliação é mista e diferentes pesos são atribuídos às dimensões avaliadas.

Na Figura 18 é possível identificar duas situações. Para valores de probabilidade superiores a 50% no caso da avaliação mista os resultados mostram que a probabilidade de um candidato ser selecionado em primeiro lugar pode aumentar consideravelmente se a avaliação apenas é baseada no desempenho científico. Considere-se os dois casos extremos, isto é, um concurso em que o desempenho científico contribui para 70% do total da avaliação final e um outro em que apenas contribui em 45%. Para estas situações verificou-se aumentos no valor da probabilidade que se situam nos intervalos $[0,2\%;12,2\%]$ e $[0,2\%;29,9\%]$ respetivamente. Para valores de probabilidade inferiores a 50% (avaliação mista) a probabilidade de um candidato ser selecionado se apenas o desempenho científico é considerado é inferior. Este comportamento está de acordo com o esperado pois o desempenho do candidato na vertente científica é baixo e portanto os parâmetros relacionados com as atividades pedagógicas, de transferência de tecnologia e de gestão serão determinantes na avaliação final do candidato, e são agora considerados na expressão que define a utilidade.

Com a informação disponível na Figura 18 é possível determinar novas funções distribuição de probabilidade para cada modelo se apenas o desempenho científico for avaliado nos concursos. Isto implica a determinação de novas probabilidades de sucesso para cada par disponível na amostra utilizada. Uma vez determinadas as probabilidades e utilizando o método de Monte Carlo calcularam-se as novas distribuições de probabilidades (Figura 19).

Uma vez disponíveis as novas funções distribuição de probabilidades torna-se viável quantificar o aumento do poder de previsão de cada um dos modelos.

Os resultados obtidos para cada modelo não diferem consideravelmente entre si. Para o modelo 1 a proporção média esperada de pares previstos corretamente é de $68\% \pm 2\%$. Para o modelo 2 e 3 é esperado em média que os mesmos prevejam corretamente em $69\% \pm 2\%$ dos

casos as decisões do painel. Os resultados sugerem que se apenas o desempenho científico fosse avaliado a previsão fornecida pelos modelos seria incrementada até 8% relativamente aos valores médios esperados para a situação inicial.

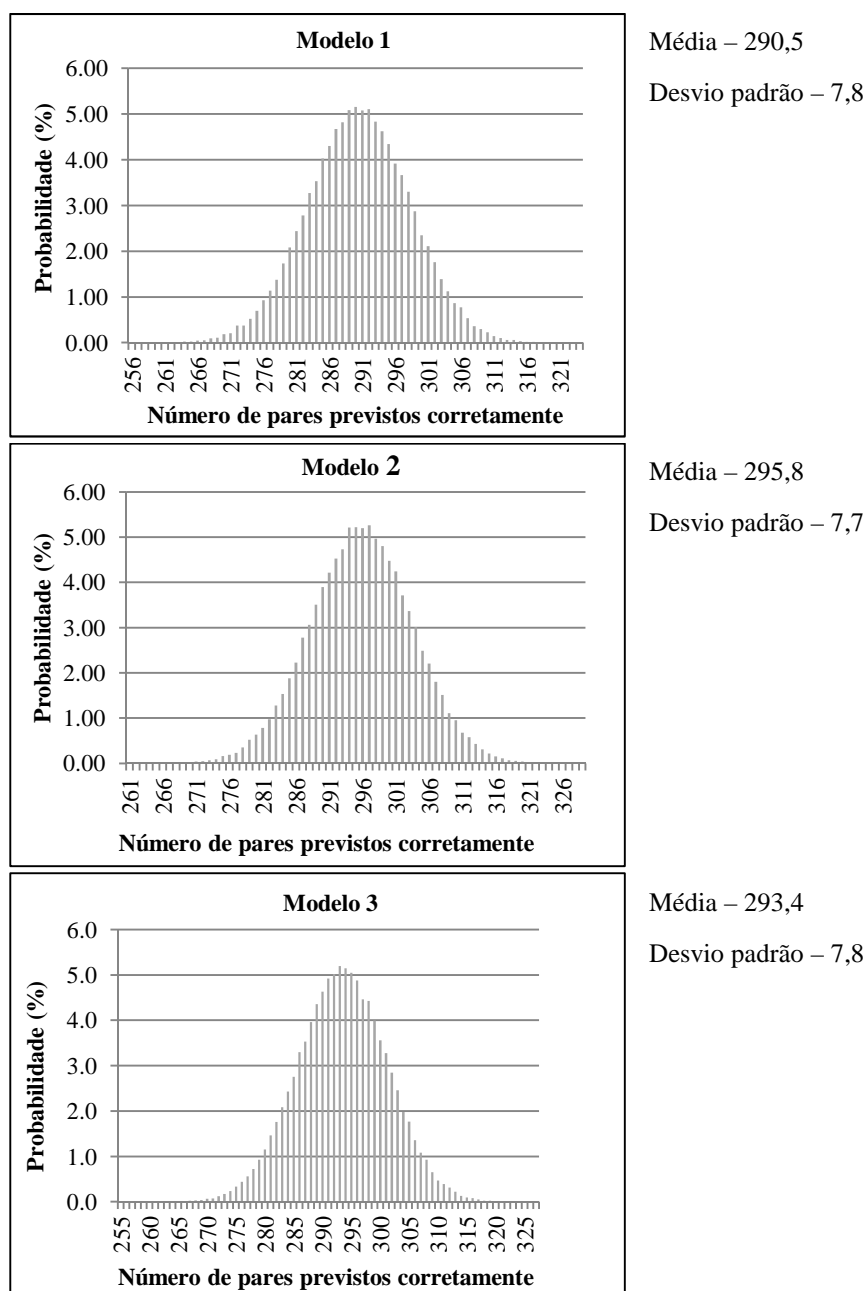


Figura 19. Função distribuição de probabilidades para cada modelo considerando que apenas o desempenho científico é avaliado.

Para obter os valores representados na Figura 19 foi necessário estabelecer um conjunto de hipóteses de trabalho (descrito na secção destinada à metodologia do presente capítulo) uma vez que nada se sabe acerca dos restantes parâmetros avaliados por cada painel. No entanto com a informação disponível é possível determinar os valores relativos ao desvio padrão para a componente que representa os parâmetros avaliados pelo painel (V_2), e que não foram considerados na definição do modelo inicial, utilizando os dados empíricos. A comparação deste valor com o valor determinado utilizando uma abordagem teórica permite

inferir sobre o rigor dos valores estimados “teoricamente”. Na Tabela 56 são apresentados os valores obtidos para o σ_{V_2} .

Tabela 56. Valores estimados para a dispersão de V_2 .

Modelo	σ_{V_2}
Modelo 1	1,09
Modelo 2	1,23
Modelo 3	1,72

Como é possível constatar os valores de σ_{V_2} estão dentro do intervalo estimado, $\left[\frac{\pi}{\sqrt{6}}; \frac{\pi}{\sqrt{3}}\right]$, com exceção do desvio padrão encontrado para o modelo 1. Perante estes resultados é possível concluir que o valor médio esperado para o número de pares previstos corretamente determinados na simulação estará muito próximo do esperado com a aplicação dos modelos definidos. De notar que os valores para a probabilidade apresentadas na Figura 18 (eixo das abcissas) resulta da integração da expressão 104 utilizando o método de Monte Carlo e assumindo que a componente V_2 segue uma distribuição Normal com média igual a zero e desvio padrão igual a $\frac{\pi}{\sqrt{6}}$.

5.7 Conclusões

Ao longo deste capítulo foi discutido o poder de previsão de cada um dos modelos considerando várias situações. Os resultados obtidos mostraram que os três modelos fornecem previsões muito próximas entre si apesar de serem baseados em diferentes indicadores.

Considerando apenas o candidato colocado na primeira posição em cada concurso, foi possível verificar que, recorrendo aos modelos desenvolvidos, a percentagem de concursos em que a probabilidade de ser selecionado em primeiro lugar é superior à disponível em um cenário aleatório é superior a 50%. De facto, os valores obtidos são razoáveis se se tiver em consideração que apenas o desempenho científico foi tomado em conta na definição dos modelos. Para a situação em que se avaliou o mesmo mas se utilizou o conjunto de pares disponíveis encontrou-se, igualmente, valores elevados. Para cerca de 75% dos pares a probabilidade de o candidato colocado na melhor posição pelo painel ser selecionado em primeiro lugar resultou superior a 50%.

Utilizando os modelos definidos espera-se que, em média, a previsão coincida com a decisão dos painéis de avaliação para mais de 60% dos casos (entre $63\% \pm 2\%$ e $65\% \pm 2\%$). Por outro lado, a probabilidade de as previsões disponíveis num cenário aleatório serem superiores àquelas que se dispõem com os modelos alcançados é inferior a 0,01%.

O poder de previsão dos modelos seria incrementado até 8% se apenas o desempenho científico fosse avaliado neste tipo de concursos. No entanto, considera-se que conclusões mais precisas apenas podem ser obtidas depois de um estudo semelhante realizado com concursos destinados exclusivamente à avaliação do desempenho científico dos investigadores.

As previsões obtidas são consideradas muito satisfatórias apesar de todas as limitações inerentes à avaliação por pares. Os modelos foram construídos com base nas decisões tomadas pelos painéis de avaliação de cada concurso. A avaliação por pares é uma metodologia afetada pela falta de consistência na opinião entre os avaliadores perante a apreciação do mesmo sujeito. Por outro lado, as características do avaliado e dos próprios avaliadores podem introduzir viés nas decisões. O conflito de interesses é um aspeto que está frequentemente presente na avaliação por pares e que pode influenciar no sentido negativo as avaliações. A presença destes aspetos na metodologia de avaliação poderá ter contribuído para diminuir o poder de previsão dos modelos. O ideal seria testar a aplicabilidade dos indicadores bibliométricos na avaliação do desempenho científico utilizando uma metodologia que fosse imune a fatores que o analista não consegue observar, o que na realidade não existe!

Apesar de todas as críticas, a avaliação por pares é utilizada desde há muitos anos e é a única metodologia disponível e aceite pela comunidade científica para a avaliação da produção científica. Sem outras opções para validar o uso de indicadores bibliométricos na avaliação do desempenho científico de indivíduos, o procedimento aqui sugerido parece ser o que melhor se adequa ao fim em vista.

Dos três modelos desenvolvidos, o modelo 1 é provavelmente aquele que será preterido. Exige um conjunto maior de indicadores bibliométricos e, no entanto, o seu poder de previsão é muito próximo do dos restantes modelos. Para além disto, como se observou no Capítulo 4,

os valores de AIC sugerem que o modelo 1 tem associada a menor probabilidade de ser o melhor na explicação das decisões finais dos painéis de avaliação. O modelo 2 e 3 apresentam o mesmo poder de previsão, mas os indicadores que os definem apresentam diferentes dimensões do desempenho científico. Cabe aqui ao painel de avaliação decidir quais os aspetos do desempenho científico que são mais relevantes.

Capítulo 6

Conclusões gerais

O estudo apresentado ao longo desta tese mostra que os indicadores bibliométricos podem constituir um instrumento de apoio à decisão final na avaliação por pares de concursos para Professor Associado e Catedrático nas universidades portuguesas.

Partindo de um conjunto de doze indicadores iniciais que se considerou descreverem as várias dimensões do desempenho científico, conseguiram-se obter três modelos que reproduzem razoavelmente, e de forma fiável e significativa, as decisões finais dos painéis de avaliação dos concursos aqui utilizados.

Apesar dos resultados bastante satisfatórios, está-se plenamente consciente de que as decisões humanas e especialmente as avaliações por pares do mérito científico de indivíduos, não podem ser reduzidas a um pequeno algoritmo matemático onde apenas se “prime numa tecla” e em segundos é providenciado o resultado. No entanto, dado o elevado esforço e custos associados à avaliação por pares, é importante encontrar um modelo matemático baseado em indicadores bibliométricos que pode ajudar na preparação das decisões finais da avaliação por pares (e que não podem em caso algum ser descartadas).

O estudo das bases de dados multidisciplinares que apresentam instrumentos que as tornam atrativas para a elaboração de análises bibliométricas evidenciou que possuem limitações relativamente à indexação. A revisão da literatura e o estudo efetuado permitem concluir que estas bases de dados não são as mais indicadas para análises bibliométricas nas áreas das ciências dos computadores, algumas engenharias, ciências sociais e humanidades.

Ao longo dos últimos anos assistiu-se a uma grande proliferação de indicadores bibliométricos. A existência de um número elevado de indicadores obrigou que se seleccionassem apenas alguns que se acredita descreverem de forma razoável a produção científica de indivíduos. Escolheram-se indicadores que permitem a comparação de indivíduos que trabalham em diferentes áreas científicas e que descrevem diferentes aspetos do desempenho científico. Constatou-se que para que os indicadores possam ser utilizados na comparação de indivíduos que pertencem a diferentes áreas de investigação é fundamental considerar as culturas de publicação e citação. Na caracterização das culturas de citação das diferentes áreas concluiu-se que o tipo de documento, o comprimento da janela de citação e a metodologia utilizada na contagem das citações são características importantes. Na caracterização das culturas de publicação o tipo de documento e o número de co-autores são parâmetros a considerar.

Nos concursos destinados à seleção de indivíduos para posições académicas em universidade devem ser sempre mencionados os aspetos a ser avaliados e o peso que cada um deve ter na decisão final. Os concursos aqui utilizados possuíam esta informação no edital publicado. Quando análises bibliométricas são requeridas pelos painéis de avaliação deve ser mencionado explicitamente quais as dimensões que se pretendem descrever. Numa situação deste género os responsáveis pela determinação dos indicadores têm a obrigação de informar os usuários finais sobre as fontes utilizadas na construção dos mesmos, das limitações apresentadas, do conjunto de pressupostos que são efetuados, das incertezas que estão envolvidas e das dificuldades que poderão surgir no momento da interpretação dos resultados.

A disponibilização deste tipo de informação permite ao utilizador final atribuir o devido valor aos indicadores e obter informação acerca das potencialidades e limitações. Isto permitirá o uso dos indicadores bibliométricos de uma forma responsável.

Existem situações para as quais, de facto, uma análise bibliométrica não permite inferir sobre o desempenho de científico de um indivíduo. Considere-se, por exemplo, um investigador a desenvolver atividades numa área emergente. Nos primeiros anos as suas atividades científicas poderão ter pouco impacto na comunidade, mas este poderá receber reconhecimento da comunidade depois de a área emergente ser aceite pela comunidade científica. Os resultados da aplicação dos modelos aqui definidos refletirão um baixo desempenho científico se avaliado na fase em que a sua investigação ainda está não foi devidamente reconhecida pela comunidade científica. No entanto um painel de avaliadores especializado tem competências que lhes permitem identificar estes casos especiais.

É importante considerar que o desempenho individual não pode ser reduzido a um simples “ número”, pois este é influenciado por muitos fatores tais como a idade do investigador, a posição no grupo de trabalho em que se insere e o domínio de investigação. Por exemplo, dentro do mesmo domínio académico e para investigadores que ocupam a mesma posição, o perfil de atividade pode diferir consideravelmente. Existirão certamente investigadores que trabalham sempre sozinhos, outros pertencerão a grupos estáveis e existirão ainda outros que colaboram ocasionalmente. Em matéria de colaboração é importante analisar se a mesma é efetuada com investigadores juniores ou seniores. Disto resulta que os indicadores bibliométricos quando utilizados para comparar investigadores têm que levar em consideração o contexto de trabalho dos indivíduos. A aplicação de indicadores bibliométricos na avaliação do desempenho a este nível implica a combinação dos mesmos com os métodos qualitativos tradicionais (avaliação por pares). Se grandes discrepâncias entre os dois métodos forem encontradas então torna-se indispensável investigar e perceber as possíveis razões para o sucedido. Em conjunto, isto irá enriquecer e melhorar o processo de avaliação.

Linhas de investigação futura

Como linhas de investigação futura propõem-se o estudo do poder de previsão dos modelos obtidos nas seguintes situações:

- 1) Um novo conjunto de concursos. O poder de previsão dos modelos foi efetuado sobre o conjunto de modelos utilizados na parametrização e o estudo deste utilizando um outro conjunto de concursos contribuirá para refinar as conclusões retiradas acerca da aplicabilidade de indicadores bibliométricos como instrumento de apoio à decisão.
- 2) Um conjunto de concursos de apenas uma área científica. Os modelos foram obtidos utilizando um conjunto de concursos de diversas áreas e para dois tipos de posições académicas. Os requisitos relativamente ao desempenho de um investigador serão, certamente, diferentes de área para área. Se numa área existe uma maior necessidade de colaboração para desenvolver um determinado trabalho, outras existirão onde este procedimento não é necessário. Numa situação deste género um indicador que fornece informação sobre colaboração poderá ser importante para a área onde a prática de colaboração é mais comum do que para a área onde esta atividade é praticamente inexistente.

3) Concursos onde o desempenho científico é o único aspeto avaliado.

Isto permitirá estudar a robustez dos modelos obtidos quando aplicados em cenários diferentes dos utilizados para a parametrização.

Seria igualmente importante estudar se os modelos obtidos resultariam os mesmos em cada uma das situações enumeradas anteriormente. As variáveis identificadas como significativas no presente estudo seriam as mesmas? Esta questão fica em aberto para o futuro.

Capítulo 7

Lista de publicações

Do trabalho desenvolvido ao longo desta tese resultaram cinco artigos. Deste conjunto quatro encontram-se publicados em revistas com circulação internacional e com avaliação por pares e um está submetido. Para além dos artigos resultaram ainda dois *proceeding papers* e quatro comunicações em forma de painel nas conferências internacionais de mais prestígio na área. Os artigos aceites e os *proceeding papers* estão indexados no *Web of Science*.

Artigos

Vieira E.S.; Cabral, J.A.S.; Gomes J.A.N.F. (2013) “How good is a model based on bibliometric indicators in predicting the final decisions made by peers? *Journal of Informetrics*. Submetido.

Vieira E.S.; Cabral, J.A.S.; Gomes J.A.N.F. (2013) “Definition of a model based on bibliometric indicators for assessing applicants to academic positions”. *Journal of the American Society for Information Science and Technology*. Aceite.

Vieira E.S.; Gomes J.A.N.F. (2011) An impact indicator for researchers, *Scientometrics*, 89(2). 607-629.

Vieira E.S.; Gomes J.A.N.F. (2011) “The journal relative impact: an indicator for journal assessment”. *Scientometrics*, 89(2). 631-651.

Vieira E.S.; Gomes JANF. (2009) “A comparison of Scopus and Web of Science for a typical university”. *Scientometrics*, 81(2), 587-600.

Proceeding papers

Vieira E.S.; Gomes J.A.N.F. (2009). “Research Quality Indicators for Brazilian, Portuguese and Spanish Universities”. 12th International Conference of the International-Society-for-Scientometrics-and-Informetrics, JUL 14-17, 2009 Rio de Janeiro. *Proceedings of the International Conference on Scientometrics and Informetrics Volume 2*, 1000-1001.

Vieira E.S.; Gomes J.A.N.F. (2009). “How to improve the citation impact of a paper: choice of journal, co-authors and institutional addresses”. 12th International Conference of the International-Society-for-Scientometrics-and-Informetrics, JUL 14-17, 2009 Rio de Janeiro. *Proceedings of the International Conference on Scientometrics and Informetrics Volume 1*, 31-39.

Comunicações em forma de painel

Vieira E.S.; Cabral, J.A.S.; Gomes J.A.N.F. (2013). “A model based on bibliometric indicators: the predictive power”. 14th International Conference of the International-Society-for-Scientometrics-and-Informetrics, JUL 15-19, 2013 Vienna, Austria. *Proceedings of the*

International Conference on Scientometrics and Informetrics Volume 2, 2054-2056. Será indexado no *Web of Science*.

Vieira E.S.; Cabral, J.A.S; Gomes J.A.N.F. (2012). “Bibliometric Indicators at individual assessment”. 17th International Conference on Science and Technology Indicators, September, 5-8, Montréal, Canada. Proceedings of Science and Technology Indicators, Volume 2, 899-900.

Vieira E.S.; Cabral, J.A.S; Gomes J.A.N.F. (2011). “The impact observed for a researcher: the contribution of several factors”. Science and Technology Indicators Conference 2011 on Actors and networks in European Science, September, 7-9, Rome, Italy

Vieira E.S.; Gomes J.A.N.F. (2010). “The journal relative impact: an index impact for scientific journals”. 11th International Conference on Science and Technology Indicators, September, 9-11, Leiden, Netherlands.

Capítulo 8

Referências

- Abramo, G. D'Angelo, C. A. e Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy*, 38(1), 206-215.
- Academy, F. (2007). Computer Science Research in Finland 2000-2006.
- Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235-246.
- Aksnes, D. W. e Taxt, R. E. (2004). Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. *Research Evaluation*, 13(1), 33-41.
- Alonso, S. Cabrerizo, F. J. Herrera-Viedma, E. e Herrera, F. (2010). hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices. *Scientometrics*, 82(2), 391-400.
- Bakkalbasi, N. Bauer, K. Glover, J. e Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(7). doi:10.1186/1742-5581-3-7
- Bar-Ilan, J. (2008). Which h-index? - A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Bar-Ilan, J. Levene, M. e Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics*, 1(1), 26-34.
- Barnett, V. e Lewis, T. (1994). Outliers in statistical data. 3 rd ed.: Wiley.
- Batista, P. D. Campiteli, M. G. Kinouchi, O. e Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179-189.
- Bettencourt, L. M. A. Kaiser, D. I. e Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics*, 3(3), 210-221.
- Board, N. S. (2012). Science and Engineering Indicators 2012. Arlington VA:National Science Foundation (NSB 12-01).
- Bonzi, S. e Snyder, H. W. (1991). Motivations for citation - a comparison of self-citation and citation to others. *Scientometrics*, 21(2), 245-254.
- Bornmann, L. e Daniel, H. D. (2005). Selection of research fellowship recipients by committee peer review. Reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics*, 63(2), 297-320.
- Bornmann, L. e Daniel, H. D. (2006). Selecting scientific excellence through committee peer review - A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, 68(3), 427-440.
- Bornmann, L. Leydesdorff, L. e Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of Informetrics*, 7(1), 158-165.
- Bornmann, L. Wallon, G. e Ledin, A. (2008). Does the Committee Peer Review Select the Best Applicants for Funding? An Investigation of the Selection Process for Two European Molecular Biology Organization Programmes. *Plos One*, 3(10).
- Box, G. E. P. Hunter, J. S. e Hunter, W. G. (2005). Statistics for experimenters: Design, Innovation, and Discovery. 2nd ed.: Wiley-Interscience.
- Burrell, Q. e Rousseau, R. (1995). Fractional counts for authorship attribution - a numerical study. *Journal of the American Society for Information Science*, 46(2), 97-102.
- Butler, L. (2008). ICT assessment: Moving beyond journal outputs. *Scientometrics*, 74(1), 39-55.
- Chapman, R. G. e Staelin, R. (1982). Exploring rank ordered choice data within the stochastic utility model. *Journal of Marketing Research*, 19(3), 288-301.
- Cicchetti, D. V. (1991). The reliability of peer-review for manuscript and grant submissions - a cross disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119-134.
- Cole, J. R. e Cole, S. (1973). Social stratification in science Chicago:The University of Chicago Press.

- Costa, C. e Oliveira, M. D. (2012). A multicriteria decision analysis model for faculty evaluation. *Omega-International Journal of Management Science*, 40(4), 424-436.
- Costas, R. e Bordons, M. (2005). Bibliometric indicators at the micro-level: Some results in the area of natural resources at the Spanish CSIC. *Research Evaluation*, 14(2), 110-120.
- Costas, R. e Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1(3), 193-203.
- Costas, R. van Leeuwen, T. N. e Bordons, M. (2010). A Bibliometric Classificatory Approach for the Study and Assessment of Research Performance at the Individual Level: The Effects of Age on Productivity and Impact. *Journal of the American Society for Information Science and Technology*, 61(8), 1564-1581.
- De Sutter, B. e Van den Oord, A. (2012). To Be or Not To Be Cited in Computer Science. *Communications of the Acm*, 55(8), 69-75.
- Delgado Lopez-Cozar, E. Robinson-Garcia, N. e Torres Salinas, D. (2012). Manipulating Google Scholar Citations and Google Scholar Metrics: simple, easy and tempting. *EC3 Working Papers 6*: 29 May, 2012.
- DGEEC (2012). *Produção Científica Portuguesa 1981-2011: Indicadores Bibliométricos*. (pp. 4).
- Donovan, J. M. e Watson, C. A. (2011). Citation Advantage of Open Access Legal Scholarship. *Law Library Journal*, 103(4), 553-573.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131-152.
- Field, A. (2005). *Discovering Statistics Using SPSS*. 2nd edition ed.: SAGE Publications.
- Fok, D. Paap, R. e Van Dijk, B. (2012). A rank-ordered logit model with unobserved heterogeneity in ranking capabilities *Journal of Applied Econometrics*, 27(5), 831-846.
- Franceschet, M. e Costantini, A. (2011). The first Italian research assessment exercise: A bibliometric perspective. *Journal of Informetrics*, 5(2), 275-291.
- Frandsen, T. F. e Nicolaisen, J. (2010). What is in a name? Credit assignment practices in different disciplines. *Journal of Informetrics*, 4(4), 608-617.
- Garfield, E. e Welljamsdorof, A. (1992). Of nobel class - a citation perspective on high-impact research authors. *Theoretical Medicine*, 13(2), 117-135.
- Gavel, Y. e Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online Information Review*, 32(1), 8-21.
- Glanzel, W. (2004). Towards a model for diachronous and synchronous citation analyses. *Scientometrics*, 60(3), 511-522.
- Glanzel, W. Schlemmer, B. Schubert, A. e Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, 68(3), 457-473.
- Glanzel, W. e Schoepflin, U. (1995). A bibliometric study on aging and reception processes of scientific literature. *Journal of Information Science*, 21(1), 37-53.
- Glanzel, W. Schubert, A. Thijs, B. e Debackere, K. (2011). A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. *Scientometrics*, 87(2), 415-424.
- Glanzel, W. e Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3), 395-404.
- Glanzel, W. Thijs, B. Schubert, A. e Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78(1), 165-188.
- Gonzalez-Pereira, B. Guerrero-Bote, V. P. e Moya-Anegon, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*, 4(3), 379-391.

- Gross, P. L. K. e Gross, E. M. (1927). College libraries and chemical education. *Science*, 66(1713), 385-389.
- Haeffner-Cavaillon, N. e Graillot-Gak, C. (2009). The use of bibliometric indicators to help peer-review assessment. *Archivum Immunologiae Et Therapiae Experimentalis*, 57(1), 33-38.
- Hagen, N. T. (2008). Harmonic Allocation of Authorship Credit: Source-Level Correction of Bibliometric Bias Assures Accurate Publication and Citation Analysis. *Plos One*, 3(12).
- Hair, J. F. B., W.C. Babin, B.J. Anderson, R.E. (2010). *Multivariate Data Analysis*. 7th Edition ed.: Prentice Hall.
- Hausman, J. e McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52(5), 1219-1240.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2), 193-215.
- Hicks, D. (2005). The Four Literatures of Social Sciences. *Journal of Management and Social Sciences*, 1(1).
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569-16572.
- Hodge, S. E. e Greenberg, D. A. (1981). Publication credit. *Science*, 213(4511), 950-950.
- Hodgson, C. (1997). How reliable is peer review? An examination of operating grant proposals simultaneously submitted to two similar peer review systems. *Journal of Clinical Epidemiology*, 50(11), 1189-1195.
- Holmes, A. e Oppenheim, C. (2001). Use of citation analysis to predict the outcome of the 2001 RAE for Unit of Assessment 61: Library and Information Management. *Information Research*, 6(2).
- Ingwersen, P. Larsen, B. Rousseau, R. e Russell, J. (2001). The publication-citation matrix and its derived quantities. *Chinese Science Bulletin*, 46(6), 524-528.
- Ingwersen, P. Larsen, B. e Wormell, I. (2000). Applying diachronic citation analysis to research program evaluations. *Web of Knowledge - a Festschrift in Honor of Eugene Garfield*, 373-387.
- Jacso, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review*, 30(3), 297-309.
- Jacso, P. (2008). The pros and cons of computing the h-index using Google Scholar. *Online Information Review*, 32(3), 437-452.
- Jayasinghe, U. W. Marsh, H. W. e Bond, N. (2001). Peer review in the funding of research in higher education: The Australian experience. *Educational Evaluation and Policy Analysis*, 23(4), 343-364.
- Jayasinghe, U. W. Marsh, H. W. e Bond, N. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: the effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society Series a-Statistics in Society*, 166, 279-300.
- Jayasinghe, U. W. Marsh, H. W. e Bond, N. (2006). A new reader trial approach to peer review in funding research grants: An Australian experiment. *Scientometrics*, 69(3), 591-606.
- Jin, B. H. Liang, L. M. Rousseau, R. e Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855-863.
- Kaiser, H. F. (1960). The application of electronic-computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- King, D. A. (2004). The scientific impact of nations. *Nature*, 430(6997), 311-316.

- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1), 35-48.
- Lariviere, V. Archambault, E. Gingras, Y. e Vignola-Gagne, E. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997-1004.
- Lariviere, V. Gingras, Y. e Archambault, E. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, 68(3), 519-533.
- Larsen, P. O. e von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575-603.
- Lewison, G. (1998). Gastroenterology research in the United Kingdom: funding sources and impact. *Gut*, 43(2), 288-293.
- Leydesdorff, L. (2003). Can networks of journal-journal citations be used as indicators of change in the social sciences? *Journal of Documentation*, 59(1), 84-104.
- Leydesdorff, L. e Opthof, T. (2010). Scopus's Source Normalized Impact per Paper (SNIP) Versus a Journal Impact Factor Based on Fractional Counting of Citations. *Journal of the American Society for Information Science and Technology*, 61(11), 2365-2369.
- Lisee, C. Lariviere, V. e Archambault, E. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology*, 59(11), 1776-1784.
- Lopez-Illescas, C. de Moya-Anegon, F. e Moed, H. F. (2008). Coverage and citation impact of oncological journals in the Web of Science and Scopus. *Journal of Informetrics*, 2(4), 304-316.
- Maroco, J. (2007). *Análise Estatística - Com utilização do SPSS*. 3rd ed.: Edições Sílabo, LDA.
- Marsh, H. W. Bonds, N. W. e Jayasinghe, U. W. (2007). Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. *Australian Psychologist*, 42(1), 33-38.
- Marsh, H. W. Jayasinghe, U. W. e Bond, N. W. (2008). Improving the peer-review process for grant applications - Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160-168.
- McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*. *Frontiers in Econometrics*: Academic Press, New York.
- Meho, L. I. e Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 59(11), 1711-1726.
- Meho, L. I. e Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- Meyer, B. Choppy, C. Staunstrup, J. e van Leeuwen, J. (2009). Research Evaluation for Computer Science. *Communications of the Acm*, 52(4), 31-34.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. *Information Science and Knowledge Management*, Vol. 9. Netherlands: Springer
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265-277.
- Moed, H. F. Debruin, R. E. e van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance - database description, overview of indicators and first applications. *Scientometrics*, 33(3), 381-422.

- Moed, H. F. e Visser, M. S. (2007). Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study. Research Report to the Council for Physical Sciences of the Netherlands Organisation for Scientific Research.
- Molinari, J. F. e Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, 75(1), 163-174.
- Montesi, M. e Owen, J. M. (2008). From conference to journal publication: How conference papers in software engineering are extended for publication in journals. *Journal of the American Society for Information Science and Technology*, 59(5), 816-829.
- Nederhof, A. J. e van Raan, A. F. J. (1987). Peer -review and bibliometric indicators of scientific performance - a comparison of *cum laude* doctorates with ordinary doctorates in physics. *Scientometrics*, 11(5-6), 333-350.
- Norris, M. e Oppenheim, C. (2003). Citation counts and the Research Assessment Exercise V - Archaeology and the 2001 RAE. *Journal of Documentation*, 59(6), 709-730.
- Norris, M. e Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2), 161-169.
- OECD (2010). Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings. (December 2010 ed.): OECD Publishing.
- Oppenheim, C. (1995). The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British -library and information-science university departments. *Journal of Documentation*, 51(1), 18-27.
- Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*, 53(5), 477-487.
- Persson, O. Glanzel, W. e Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432.
- Pestana, M. H. e Gageiro, J. N. (1998). Análise de dados para ciências sociais -A complementariedade do SPSS. 1ª ed. Lisboa: Edições Sílabo, Lda.
- Pinski, G. e Narin, F. (1976). Citation influence for journal aggregates of scientific publications - theory, with application to literature of physics. *Information Processing & Management*, 12(5), 297-312.
- Prathap, G. (2010). Is there a place for a mock h-index? *Scientometrics*, 84(1), 153-165.
- Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, 25(4), 348-&.
- Reale, E. Barbara, A. e Costantini, A. (2007). Peer review for the evaluation of academic research: lessons from the Italian experience. *Research Evaluation*, 16(3), 216-228.
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4(2), 131-134.
- Reuters, T. (2012a). Journal citation reports. http://admin-apps.webofknowledge.com/JCR/help/h_using.htm#UsingtheJCRWisely. Consultado em Novembro 2012.
- Reuters, T. (2012b). Methodolgy for ScienceWatch.com. <http://archive.sciencewatch.com/about/met/>. Consultado em Novembro 2012.
- Reuters, T. (2012c). The Thomson Reuters Journal Selection Process http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/. Consultado em Novembro 2012.
- Rinia, E. J. van Leeuwen, T. N. van Vuren, H. G. e van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria - Evaluation of condensed matter physics in the Netherlands. *Research Policy*, 27(1), 95-107.

- Rothwell, P. M. e Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience - Is agreement between reviewers any greater than would be expected by chance alone? *Brain*, 123, 1964-1969.
- Schubert, A. e Braun, T. (1986). Relative indicators and relational charts for comparative-assessment of publication output and citation impact. *Scientometrics*, 9(5-6), 281-291.
- Schubert, A. e Glanzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. *Journal of Informetrics*, 1(3), 179-184.
- Schubert, A. Glanzel, W. e Braun, T. (1986). Relative indicators of publication output and citation impact of european physics research, 1978-1980. *Czechoslovak Journal of Physics*, 36(1), 126-129.
- SCImago (2011). SIR- SCImago Institutions Rankings.
<http://www.scimagoir.com/index.php#>. Consultado em Novembro, 2011.
- Scopus (2011). Content Coverage Guide.
http://www.info.sciverse.com/UserFiles/sciverse_scopus_content_coverage_0.pdf. Consultado em November 2012.
- Scopus (2013a). Journal M3trics. <http://www.journalmetrics.com/about-journal-metrics.php>. Consultado em Janeiro 2013.
- Scopus (2013b). Scopus Content Coverage Guide.
<http://files.sciverse.com/documents/pdf/ContentCoverageGuide-jan-2013.pdf>. Consultado em Janeiro 2013.
- Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, 43(9), 628-638.
- Shapiro, S. S. e Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). [Article]. *Biometrika*, 52, 591-&.
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Second ed., Vol. 2009: Cambridge University Press.
- van Eck, N. J. e Waltman, L. (2012). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison arXiv:1208.6122.
- van Hooydonk, G. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, 48(10), 944-945.
- van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics*, 59(3), 467-472.
- van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502.
- van Wijk, E. e Visser, M. S. (2012). *Bibliometric Analysis of Universidade Nova de Lisboa 2004-2010/11*. Center for Science and Technology Studies (CWTS).
- Ventura, O. N. e Mombru, A. W. (2006). Use of bibliometric information to assist research policy making. A comparison of publication and citation profiles of Full and Associate Professors at a School of Chemistry in Uruguay. *Scientometrics*, 69(2), 287-313.
- Vieira, E. S. e Gomes, J. A. N. F. (2010). Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, 4(1), 1-13.
- Vieira, E. S. e Gomes, J. A. N. F. (2011). An impact indicator for researchers. *Scientometrics*, 89(2), 607-629.
- Visser, M. S. van Raan, A. F. J. e Nederhof, A. J. (2007). *Bibliometric Benchmarking Analysis of the Universidade Nova de Lisboa 2000-2006*. Center for Science and Technology Studies (CWTS).
- Wagenmakers, E. J. e Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192-196.

- Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, 6(4), 700-711.
- Waltman, L. van Eck, N. J. van Leeuwen, T. N. e Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, 7(2), 272-285.
- Waltman, L. van Eck, N. J. van Leeuwen, T. N. Visser, M. S. e van Raan, A. F. J. (2011a). Towards a new crown indicator: an empirical analysis. *Scientometrics*, 87(3), 467-481.
- Waltman, L. van Eck, N. J. van Leeuwen, T. N. Visser, M. S. e van Raan, A. F. J. (2011b). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37-47.
- Warner, S. L. (1962). Stochastic choice of mode in urban travel : a study in binary choice. Metropolitan transportation series: Northwestern University Press.
- Wood, M. Roberts, M. e Howell, B. (2004). The reliability of peer reviews of papers on information systems. *Journal of Information Science*, 30(1), 2-11.
- Zitt, M. e Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, 59(11), 1856-1860.